DOCUMENT RESUME

ED 207 609 IR 009 728

AUTHOR	Schuler, S. C., Ed.
TITLE	Manual of Documentation Practices Applicable to Defence-Aerospace Scientific and Technical Information. Volume II: Sections 4--Data Recording and Storage; 5--Mechanization Systems and Operations; 6--Announcement Services and Publications.
INSTITUTION	North Atlantic Treaty Organization, Neuilly-sur-Seine (France). Advisory Group for Aerospace Research and Development.
REPORT NO	AGARD-AG-235; ISBN-92-835-1331-2
PUB DATE	Jul 79
NOTE	132p.; For related documents, see IR 009 727-730.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Computer Programs; *Computers; *Data Processing; Flow Charts; Information Centers; Information Processing; *Information Retrieval; *Information Services; *Information Storage; Online Systems; *Publications; Selective Dissemination of Information
IDENTIFIERS	*Scientific and Technical Information

ABSTRACT

The second of four volumes in a series describing the basic documentation practices involved in the initial setting up and subsequent operation of an information-library organization to provide defense-aerospace scientific and technical information services, this manual consists of three sections. "Data Recording and Storage," by J. Howard Petrie, provides an introduction to the hardware and software of computer systems, discusses the problems of inputting data, describes different types of input and storage equipment, and outlines management and systems analysis problems in the project environment. "Mechanization Systems and Operations," by Victor Rogers, provides a working basis for setting up a computer system for indexing, processing, and disseminating information, mainly in the form of bibliographic references. The main emphasis is on the in-house computer and methods are described for setting up, storing, and exploiting databases, and creating an announcement journal. "Announcement Services and Publications," by Elizabeth Ridler, reviews methods for announcing holdings and new acquisitions to users, provides examples of manually and computer produced bulletins, and discusses the production of indexes to computer produced publications. Also described are manual and computer-based SDI services and repackaging of literature resources as bibliographies, state of the art reports, and packaged information for technical innovation. (Author/RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED207609

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AGARD-AG-235
Volume II

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARDograph No.235

MANUAL OF DOCUMENTATION PRACTICES APPLICABLE TO DEFENCE-AEROSPACE SCIENTIFIC AND TECHNICAL INFORMATION

VOLUME II

containing

Section
- 4 – DATA RECORDING AND STORAGE
- 5 – MECHANIZATION SYSTEMS AND OPERATIONS
- 6 – ANNOUNCEMENT SERVICES AND PUBLICATIONS

(The complete Publication Layout appears on page iv.)

IR009728

This AGARDograph has been prepared at the request of the Technical Information Panel of AGARD.

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

Published July 1979

Copyright © AGARD 1979
All Rights Reserved

ISBN 92-835-1331-2

Set and printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

GENERAL FOREWORD

The purpose of this Manual is to describe in a series of separately-published Volumes the basic documentation practices which are involved in the initial setting up, and subsequent operation of an Information-Library Organisation to provide defence-aerospace scientific and technical information services.

- The manual is primarily intended for the main defence-aerospace information centres in the smaller nations, and the specialised defence establishments and defence contractors in the larger NATO countries.
- For those information centres which already have a well-developed system, the manual may prove helpful in the work of analyzing and evaluating existing system performance, or in revising an ineffective system. An important subsidiary objective is therefore to encourage the greater use of modern techniques of information processing.
- The manual endeavours to meet the needs of a wide spectrum of readers – the senior man concerned with setting up a new system, as well as junior staff who may be using the manual as a training aid.
- The various Sections aim to focus on the problems and techniques associated with processing unpublished reports and related information, rather than conventional book-journal libraries. Emphasis is on practical solutions and, where appropriate, useful operating suggestions.

The manual has been planned by the AGARD Technical Information Panel and will consist of four Volumes comprising twelve Sections in all, each prepared by a well-known expert in the field. The Publication Layout is given on the following page and publication of the complete set will be spread over three to four years.

S.C.SCHULER
General Editor
(Former Chairman, AGARD
Technical Information Panel)

MANUAL OF DOCUMENTATION PRACTICES APPLICABLE TO DEFENCE-AEROSPACE SCIENTIFIC AND TECHNICAL INFORMATION

PUBLICATION LAYOUT

Section

VOLUME I

1 ACQUISITION & SOURCES, by P.F.Eckert
Types of material, screening, evaluation, sources in Governments and other countries, information on current research

2 DESCRIPTIVE CATALOGUING, by B.P.Gladd, O.G.Luchaka and J.C.Wade
Functions, standardisation, corporate authors and other compatibility factors, document process sheets, manual and computer translations

3 ABSTRACTING & SUBJECT ANALYSIS, by T.C.Bearman
Abstracting standards, descriptor allocation, thesaurus, computer-aided indexing

VOLUME II

4 DATA RECORDING & STORAGE, by J.H.Petrie
Data preparation rules, handling chemical compounds and scientific symbols, etc., card systems, tape typewriters, introduction to computerised operations

5 MECHANIZATION SYSTEMS & OPERATIONS, by V.J.Rogers
Basic mechanisation, in-house computer, minicomputers, bureau working

6 ANNOUNCEMENT SERVICES & PUBLICATIONS, by E.H.Ridler
Selective dissemination of information, bulletin production, newsletters and digests, bibliographies

VOLUME III*

7 SEARCH & INFORMATION RETRIEVAL
Manual systems, searching by computer, inverted files, KWIC/KWOC, on line systems, batch working, packaging of computer output

8 DISSEMINATION & USER NEEDS
Register of users, initial distribution, specific requests, recording and recall, guidelines on sensitive aspects, conditions of release, state of art reviews, translations

9 MICROFORM SYSTEMS & REPROGRAPHY
Microfiche preparation and duplication, COM/CIM, readers and printers, photocopying, printing processes

VOLUME IV*

10 SECURITY STORAGE, REGRADING
Security grading procedures, storage methods, weeding

11 ORGANISATION & MANAGEMENT
Aims and objectives, staffing, promotional activities, identifying users

12 NETWORKS & EXTERNAL SOURCES OF INFORMATION
National and international

---

* **To be issued later.** The contents listed here are those presently proposed: they may be amended, however, as their prepration progresses.

Section 4

DATA RECORDING AND STORAGE

by
J.Howard Petrie
Commission of the European Communities
Directorate General for Scientific
and Technical Information and
Information Management (DG XIII)
Luxembourg

ABSTRACT

An introduction to the hardware and software of computer systems is given, followed by a description of the problems of the input of data, including the representation of characters and the choice of a character set. The different types of input and storage equipment are described and examples are given. Management and systems analysis problems are outlined in the project environment. Finally, there is a brief introduction to chemical structure input and storage techniques.

CONTENTS

Page

1. INTRODUCTION 7
  1.1 Contents of the Section 7
  1.2 The Use of Computers – Some Considerations 7
  1.3 Computers for Library and Information Services 7

2. COMPUTER HARDWARE AND SOFTWARE 8
  2.1 Components of a Computer 8
  2.2 Computer Hardware 8
  2.3 Computer Software 10

3. INPUT OF DATA TO A COMPUTER 12
  3.1 The Problem 12
  3.2 Internal Computer Character Representation 12
  3.3 Choosing a Character Set 13
  3.4 Labels and Formats 14
  3.5 Accuracy 16

4. INPUT EQUIPMENT 17
  4.1 Introduction 17
  4.2 The Punched Card 17
  4.3 Punched Paper Tape 18
  4.4 Key-to-Magnetic Storage Devices 19
    4.4.1 Introduction 19
    4.4.2 Hardware 20
    4.4.3 Capabilities of Key-to-Storage Devices 21
    4.4.4 Comparison of Key-to-Storage Devices with Punched Tape and Cards 21
  4.5 On-Line Input 21
    4.5.1 General 21
    4.5.2 Software 23
  4.6 Optical Character Recognition (OCR) 24
    4.6.1 Description 24
    4.6.2 Character Sets 24
    4.6.3 Make-up of an OCR Reader 24
    4.6.4 Feasibility of the OCR Solution 24
  4.7 Choosing Input Equipment 27

5. INFORMATION STORAGE DEVICES 28
  5.1 Introduction 28
  5.2 Main Memory 28
  5.3 Direct-Access Devices 28
  5 4 Magnetic Tape 30
  5.5 Mass Storage Device 30
  5.6 Photocopy and Laser-Holographic Storage 30
  5.7 Newer Memory Types 30

6. THE MANAGEMENT OF PROJECTS 31
  6.1 The Project Environment 31
  6.2 The Systems Approach 31
    6.2.1 The Stepwise Approach to System Implementation 31
    6.2.2 Systems Analysis and Definition 31
    6.2.3 Systems Design 32
    6.2.4 Construction and Installation 32
  6.3 Personnel 33
  6.4 Other Project Management Considerations 33
  6.5 Contract Problems 34
  6.6 Operational Problems 34

7. CHEMICAL STRUCTURE INFORMATION SYSTEMS – A BRIEF INTRODUCTION 35

REFERENCES 36

ACKNOWLEDGEMENT 37

Page

APPENDICES

A. AN EXAMPLE OF ON-LINE DATA ENTRY AT THE LIBRARY OF THE INTERNATIONAL DEVELOPMENT RESEARCH CENTRE (CANADA) 38

B. SOME EXAMPLES OF INPUT FORMATS, FORMS, HANDLING OF SPECIAL SYMBOLS AND ENTRIES IN A DATA DICTIONARY 42

C. PROCESSING OF BIBLIOGRAPHIC INFORMATION AT THE CANADIAN NATIONAL DEFENCE INFORMATION CENTRE 47

LIST OF TABLES

		Page
Table 1	Some Data Verification Techniques	17
Table 2	Advantages and Disadvantages of the Punched Card as an Input Medium	18
Table 3	Advantages and Disadvantages of Paper Tape as an Input Medium	19
Table 4	Advantages and Disadvantages of Key-to-Storage Devices	23
Table 5	Steps to be Taken in the Choice of Input Equipment	27

LIST OF FIGURES

Page

Figure 1 The major hardware components of a digital computer 9

Figure 2 The components of the DIMDI relation-based data base management system 11

Figure 3 International reference version of the 7-bit coded character set for information processing interchange (ISO-646) 12

Figure 4 The Datamatic tape-typewriter keyboard character set 13

Figure 5 Representation of special symbols in DRIC data base 14

Figure 6 An example of input to SDIM 15

Figure 7 Calculation of a modulus II check digit 16

Figure 8 An example of a punched card 18

Figure 9 Punched paper tape 19

Figure 10 Schematic diagram of a DATAPOINT® key-to-cassettee tape processor 20

Figure 11 Components of a basic key-to-storage system 21

Figure 12 The NIXDORF key-to-storage input equipment 22

Figure 13 Elements of a shared processor key-to-disc system 22

Figure 14 A pre-formatted screen for the DIMDI DINUPS input system 23

Figure 15 The OCR-B character set of ISO 25

Figure 16 Input of bibliographic information to the COMPUSCAN OCR reader 26

Figure 17 Elements of an OCR reader 26

Figure 18 Outline of a magnetic drum 29

Figure 19 Section through a fixed-head magnetic disc storage unit 29

Figure 20 Section through a moving-head magnetic disc storage unit 30

1. INTRODUCTION

1.1 Contents of the Section

This Section of the Manual is designed to give an insight into the potential for using the computer for library and information service applications. In addition, organizations which already use computer systems may want to improve them and it is hoped that the Section will be of use to them, too. Hence, some subjects are treated at a superficial level and others in greater depth. It begins with a brief introduction to some of the ways in which a computer might be used in libraries and information systems. It is followed (Chapter two) by a simplified explanation of the computer, designed to give those unfamiliar with the concepts an insight into the major hardware and software points. Chapter three deals with the problem of data input, including choice of a character set, data labels, formats and verification techniques. This is followed by a review of types of input equipment designed to help in the selection of equipment (Chapter four). Chapter five contains a short introduction to computer storage devices. Microforms are not covered as this will be dealt with in Volume III of the Manual. Chapter six deals with the management aspects of the introduction and use of computer-based information systems. Finally, there is a short overview of chemical information input and storage.

An attempt has been made to keep the explanations as simple as possible and hence, in order to get a more detailed picture, the reader should consult some of the many works available, including those cited in the bibliography. For problems with terminology, the reader is advised to turn to Gibbs[1] (AGARDograph 182). The 'Encyclopaedia of Computer Science' by Ralston and Meek[2] is also a useful one-volume source of information. 'The Encyclopaedia of Library and Information Science'[3] is a comprehensive multi-volume work, useful for explaining individual points which might not be clear.

1.2 The Use of Computers – Some Considerations

During the past decade or more, the computer has played an increasingly important part in the services provided by library, documentation and information centres; its role is considered to be essential in many organizations and is likely to increase in importance as facilities become more sophisticated and equipment becomes cheaper. Even now, it is important for management to examine the potential of using the computer for solving the problems of information storage and retrieval. It will not be long before organizations which do not use computers will be at a serious disadvantage compared with more go-ahead competitors. Managers who do not realize the potential of computers as a resource will lose out to others who take advantage of them. Properly managed, the computer will provide a valuable aid to make the information service more efficient and effective, complementing and making the information staff more effective by giving them a powerful tool to assist them in their work.

'Properly managed' are important keywords, for the computer is something which cannot be approached casually. Computers have, in the past, created big problems for unprepared managers. Fortunately, the techniques of systems analysis and project management are today on a sounder footing as a result of past mistakes.

Some of the gains a computer can bring are:

- they are consistent and generally more accurate in undertaking routine tasks than human beings;
- they can undertake clerical tasks much more quickly than human beings;
- the information stored can be used flexibly. A single record can be used many times and for different applications. For instance, a record of a recently acquired book or report can be used to rapidly produce a list of new acquisitions and for a catalogue of the holdings of the library;
- machine-readable information can be taken from or given to other organizations, e.g., centrally- produced cataloguing records (MARC Machine-Readable Catalogue of the US Library of Congress) or records from an abstracting service (e.g., Chemical Abstracts Service).

There are disadvantages compared with manually-held files:

- the computer has to be told exactly how to perform its work;
- the machine sometimes breaks down;
- the files of information in the computer are not always completely up to date;
- capital outlay can be expensive;
- the adoption of a computer can mean changes in staffing.

1.3 Computers for Library and Information Services

Libraries and information services have files of information which are designed to serve particular purposes, for instance:

- a list of users,
- a catalogue of the stock of a library,
- subject indexes,
- a bibliography of articles in a particular subject field.

Computers can be used to store such information and in many cases, do so with advantages, compared with manually maintained files, such as card catalogues. Some of the ways that computers are effectively used in libraries and information services are:

- the production of abstracts journals and the associated indexes. The same information can be stored in the computer and searches can be made by putting questions directly to the computer;
- the production of library catalogues. The same information can be searched for directly in the computer (see Appendix C);
- ordering systems for library materials;
- library circulation systems;
- the storage and retrieval of scientific data e.g. the structure and function of chemical substances.

The two basic ways of using a computer are:

- batch processing,
- on-line mode.

Using batch processing, many similar units of information are collected together and are processed by the computer at the same time (in batches). This technique has the advantage of being less expensive than on-line, real-time processing, but the latest information is not available until a batch has been processed. On-line mode can be defined in different ways, but basically, a user has a visual display terminal which permits an interaction with the information stored in the computer. Some systems allow the user to retrieve information within a short time, although modifications and new information have to be entered in batches. Some other systems allow both entry and retrieval to be made immediately. When units of information are treated individually by the computer the technique is called transaction processing, i.e., the computer responds in real time within a few seconds.

The facilities which are justifiable for a particular application will depend on the needs. For instance, it is a requirement of an airline seat reservation system that a cancelled seat be immediately available and hence immediate updating is required. Increased demands lead to increased costs because the computer system needs to be more sophisticated. Because of the power of larger computers, a number of users can be served simultaneously. This can be an efficient way of using a computer system, i.e., to share its resources with other users. The main ways of acquiring computer power are:

- to purchase services from a computer bureau (this can include personnel as well as computer power);
- to use part of a computer available within the institution;
- to have a computer dedicated to the library and information service.

The increasing power and decreasing costs of minicomputers are making the last alternative more attractive. The subject has been treated in AGARD publication LS-92 (Ref.4). These problems are dealt with in more depth in Chapter six.

External computer-based services are now available which allow access, using a terminal, to computer-based information systems. Thus it is possible to search a wide range of bibliographic data bases and factual data banks to obtain the references to documents of interest and factual information. The computer may be many kilometers away from the terminal, the information being carried on communications lines, often on the normal telephone network. The user is charged for the services used. With such external services it is only necessary to acquire a terminal and learn how to use it. On-line, real-time access and the use of a terminal are at opposite extremes in terms of the commitment of the organization to the use of the computer. An information service will have to decide the level of commitment it wishes to have, by examining costs, performance and benefits.

2. COMPUTER HARDWARE AND SOFTWARE

2.1 Components of a Computer

A computer is a device in which information can be stored and organized, and from which it can be retrieved and displayed. Simply speaking, a computer consists of the following:

- hardware: the physical equipment which stores, manipulates, transmits and transforms information;
- software: the sets of instructions which tell the computer what to do.

In practice, hardware and software are two inter-related subsystems, both of which are necessary for the correct functioning of the machine.

2.2 Computer Hardware

The computer contains a number of devices which transduce (i.e. transform data/information from one physical form to another), store, transform, communicate and control. An outline of the components of a computer are shown in Figure 1. The devices which make up a computer are described in detail by Mao[5] and they have the following functions:

- storage,
- input,
- output,
- control,
- arithmetic and logic operations,
- communications.

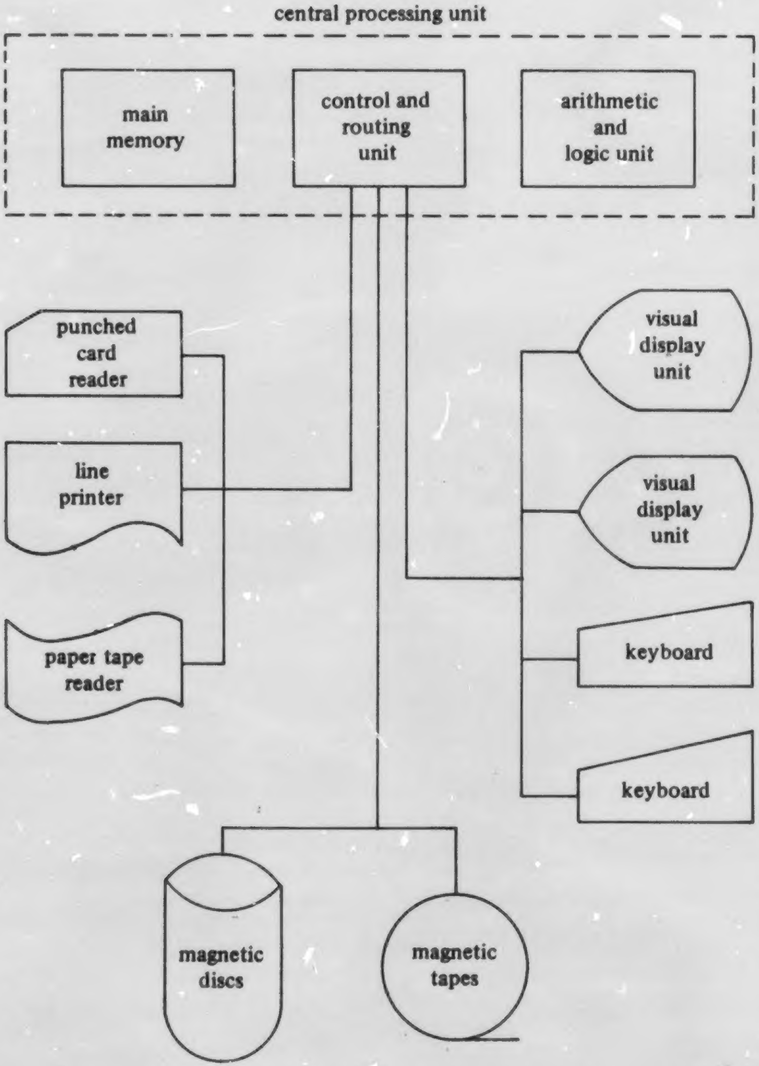

central processing unit
main memory
control and routing unit
arithmetic and logic unit
punched card reader
line printer
paper tape reader
visual display unit
visual display unit
keyboard
keyboard
magnetic discs
magnetic tapes

Fig. 1 The major hardware components of a digital computer

Storage devices hold the information that has been put into a computer. Information is stored and processed as binary numbers using 'bi-stable' devices. An electric light switch can be regarded as such a device, for it can either be off or on. If the number '1' is equivalent to 'on' and '0' to off, then a number could be stored in a series of such devices (bits). Letters of the alphabet, numbers and other symbols are all reduced to representative numbers in the computer. This is explained in more detail in chapter three. The physical characteristics of storage devices usually determine the speed with which data can be retrieved (access time), the amount of data which can be stored and the amount of data which can be transferred from the device. *Main memory* is used to store programs and the information which is currently being processed and hence the speed of access is extremely fast.

A number of techniques are used to produce main memories, the most widely known of which is the ferrite core. A store consists of a series of cores each of which can be magnetized in one of two directions to store either a '1' or a

'0'. Semiconductor memories are increasingly being used; these are binary digital memories which use an electronic circuit for each memory cell. Discs, drums and magnetic tapes are examples of storage devices which hold more information than main memory but which have a slower access time. They are often used to store files of data. Storage devices are discussed in more detail in chapter five.

Input devices are used to put the information into the computer. Punched card machines and visual display units are examples of such transducers by which information is transferred from paper into machine readable form. Each individual letter, number, symbol etc. is translated into a corresponding pattern of ones and zeros. Input devices are discussed in chapter four and the problems of putting data into the machine is described in general in chapter three.

Output devices are used to display information and a variety of machines are available. Printers with speeds ranging from tens of characters per second to 20,000 lines per minute are in use. Video display units can be used for displaying information in alphanumeric or in graphical form. Microfilm can be produced directly by COM (computer-output-microfilm) and graphs can be plotted using graph plotting equipment. Computer typesetting devices are used to produce printed texts directly from the computer.

The *control unit* issues commands and control signals to other functional units. For instance, it fetches instructions from main memory, decodes them, selects the part of the arithmetic and logic unit to carry out a calculation and transmits the data to it.

The *arithmetic and logic unit* is that part of the machine in which all arithmetic and logical operations are carried out. The arithmetic and logic unit, together with the control unit and main memory are often referred to as the central processing unit (CPU).

Communications between functional units is supervised by the control unit (see Figure 1). Peripherals (e.g., magnetic discs or visual display units) may demand attention by interrupting the CPU. Computer peripherals may be at some distance from the CPU and communicate via telecommunications lines or specially constructed networks. Sometimes a small computer (called a front-end processor) is attached to the main computer to handle video display terminals and other computer peripherals which are at a distance from the main computer.

Details of most of the hardware produced by the major manufacturers are available in the Auerbach computer technology reports.[6]

2.3 Computer Software

Computer software is the means used to tell the computer what to do. It consists of programs which comprise logical series of instructions written in a language convenient to human beings. Fortran, Cobol and PL1 are well known languages. A program is translated into the language of the machine using another program called a compiler. The translated program consists of a series of very specific instructions (e.g. add, divide, etc.) which are stored in main memory and are executed sequentially by the arithmetic and logic unit. For a more detailed description see Ralson[2].

A major part of the computer system is the operating system, a series of programs which also reside in main memory. These programs are designed to undertake the various supervisory and management tasks designed to maximize the efficiency of the computer. For instance the operating system decides where in main memory a program should reside. Operating systems vary greatly in their capabilities and are much more sophisticated for large computers than for small. For instance, some systems provide for on-line real-time working and others only for batch processing. Coffman[7] covers the subject of operating systems in detail.

Computers usually have a range of software which is designed to ease the task of programming. Examples of tasks undertaken by such supporting programs are:

- sorting of information into a specified order (sort programs),
- data base management,
- compilers.

Depending on the task in hand, some of the above programs are used by, or included alongside, the application program. Because of the increasing cost, complexity and importance of software it is often divided by the manufacturer into smaller parts which are then sold separately (unbundling).

Supporting software and packages, which can be used on their own to perform specific tasks, are available from sources other than the manufacturer of the computer system. Organizations exist which produce software products and sell them. Goetz[8] reviews the future prospects of the software products industry. Packages can be the best approach to software production because:

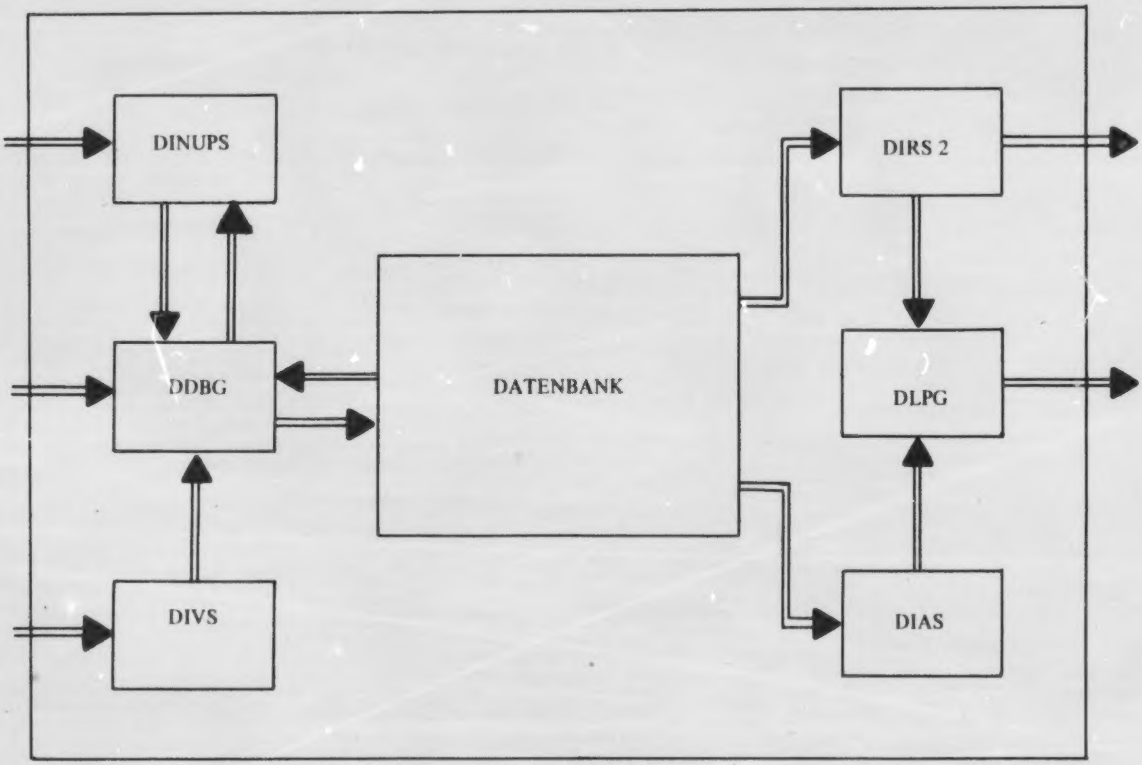
- software products are usually sold with a guarantee and are maintained;
- detailed documentation is usually available;
- new versions become better but are made compatible with older versions;

- if a package is distributed widely, the purchase price can be reasonable;
- a general package is usually reliable because it has many users.

However, software packages may not be able to cater for specific local needs.

Data base management systems are important pieces of supporting software which help in the use of the computer. Using DBMS, information can be controlled, manipulated and retrieved in an optimal fashion. DBMS are worthy of mention because the choice of the correct one can be a critical factor in determining how much effort will be needed to set up and operate computer systems. Because they are supporting software, a certain amount of programming will need to be done and that amount will depend to a large extent on the facilities provided by the DBMS. Some DBMS are more suited to bibliographic applications than others. In fact there are DBMS which have been devised especially for such applications. For instance, the GRIPS (General Relation based Information Processing System, produced by the Deutsches Institut für Medizinische Dokumentation und Information) has facilities which are specially designed for managing bibliographic data bases. An outline of the features is shown in Figure 2. GRIPS permits the user to input information on-line, to store it in the form of relations (see Reference 9) and retrieve the data with an associated information retrieval system. A report writer, which facilitates the compilation of special listings in a desired format and of statistical information is also available.

Somewhat similar is the software produced by the International Development Research Centre in Canada (IDRC)[10]. The software is more orientated towards library applications than GRIPS, but it is also built on a relational DBMS. The system is novel in that it is implemented on a minicomputer. A number of DBMS are available for minicomputers, but their capabilities need to be examined even more carefully than those for larger machines for they may offer fewer facilities. Indeed IDRC decided to build their own. A survey of commercial DBMS for minicomputers has been made by Grosch[11]. Auerbach[6] surveys DBMS in general. A number of books, for example,[12,13,14] treat the problems of data base management systems.



DINUPS: INPUT AND UPDATE SYSTEM
DDBG: DATA BASE GENERATOR
DIVS: INFORMATION VALIDING SYSTEM
DIRS 2: INFORMATION RETRIEVAL SYSTEM 2
DLPG: LIST PROGRAM GENERATOR
DIAS: ADMINISTRATION SYSTEM

Fig.2 The components of the DIMDI relation-based data base management system, GRIPS

3. INPUT OF DATA TO A COMPUTER

3.1 The Problem

In order to store information in a computer, the information has first of all to be put into the machine. This chapter deals with input and the associated problems such as making sure it is correctly recorded, and considers how the information must be represented in order to be of most use.

3.2 Internal Computer Character Representation

There are several ways of representing information in the computer and the best way depends on the nature of the information to be presented to the machine. For the most part, information for documentation is in the form of symbols i.e. letters, numbers or special characters. Computers are designed to manipulate such information; they have a 'character set' or standard set of symbols, each of which has a corresponding pattern of bits. Lynch[15] deals with the problem of character manipulation in a computer. An example of a set of characters in wide use is that adopted by the International Organization of Standardization (ISO) (Ref.16). The set of characters is shown in Figure 3. It has been adopted by many

				b_7	0	0	0	0	1	1	1	1
				b_6	0	0	1	1	0	0	1	1
				b_5	0	1	0	1	0	1	0	1
b_4	b_3	b_2	b_1	column / row	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	TC$_7$ (DLE)	SP	0	@	P	`	p
0	0	0	1	1	TC$_1$ (SOH)	DC$_1$	!	1	A	Q	a	q
0	0	1	0	2	TC$_2$ (STX)	DC$_2$	"	2	B	R	b	r
0	0	1	1	3	TC$_3$ (ETX)	DC$_3$	#	3	C	S	c	s
0	1	0	0	4	TC$_4$ (EOT)	DC$_4$	¤	4	D	T	d	t
0	1	0	1	5	TC$_5$ (ENQ)	TC$_8$ (NAK)	%	5	E	U	e	u
0	1	1	0	6	TC$_6$ (ACK)	TC$_9$ (SYN)	&	6	F	V	f	v
0	1	1	1	7	BEL	TC$_{10}$ (ETB)	'	7	G	W	g	w
1	0	0	0	8	FE$_0$ (BS)	CAN	(	8	H	X	h	x
1	0	0	1	9	FE$_1$ (HT)	EM	)	9	I	Y	i	y
1	0	1	0	10	FE$_2$ (LF)	SUB	*	:	J	Z	j	z
1	0	1	1	11	FE$_3$ (VT)	ESC	+	;	K	[	k	{
1	1	0	0	12	FE$_4$ (FF)	IS$_4$ (FS)	,	<	L	\	l	|
1	1	0	1	13	FE$_5$ (CR)	IS$_3$ (GS)	-	=	M	]	m	}
1	1	1	0	14	SO	IS$_2$ (RS)	.	>	N	^	n	‾
1	1	1	1	15	SI	IS$_1$ (US)	/	?	O	_	o	DEL

Fig.3 International reference version of the 7-bit coded character set for information processing interchange (ISO – 646)

countries for use as a national standard, the best known being the ASCII (American Standard Code for Information Interchange) character set. The information is contained in 7 bits which gives a total of 128 (2^7) different characters. For instance:

Character	*Code (bits)* *Binary*	*Decimal*
A	1000001	65
B	1000010	66
a	1100001	97
b	1100010	98
?	0111111	63
2	0110010	50

The character set used by IBM is called EBCDIC (Extended Binary Coded Decimal Interchange Code). This is an eight bit code, giving 2^8 or 256 possibilities. An 8 bit version of ASCII is also available.

A knowledge of the codes used to represent characters is important especially from the point of view of the order in which the characters appear (code sequence). For instance, in ASCII, the numbers have lower codes than letters (A=65, 2=50): the opposite is the case of EBCDIC. Since sorting of characters is performed on the value of the codes, the desired order of alphabetization may have to be imposed using special techniques.

3.3 Choosing a Character Set

The size of the character set required will be dependent on the application. For many applications, the normal ISO character set is sufficient. For other applications, e.g. for the representation of mathematical equations, a larger set will be required. The size of the internal character set is often not the limiting factor from the point of view of the computer. Input and output devices often restrict the character set further. Some devices, such as the card punch and line printer, often have upper case only.

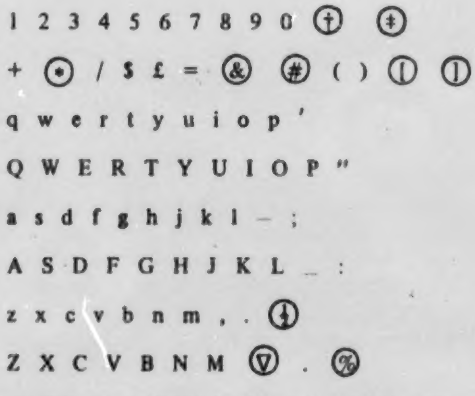
There are many applications for which it may be necessary to have a character set larger than the internal set. In computer typesetting operations, for instance, not only must a larger set be stored (e.g., Greek letters) but information on the type font (e.g., itallic, bold) must be given. In such cases, special conventions are used. For instance, for the publications of Inspec (The Institution of Electrical Engineers in the United Kingdom)

$u2 is used for superscript ^{2}
$gS is used for Greek letter Σ

Special programs have to be written to manipulate these characters.

Many information centre and library applications can be covered by the normal character set. The set used at the United Kingdom Defence Research Information Centre (DRIC) (Ref.17) is shown in Figure 4. Special character problems are mainly dealt with by spelling out those which cannot be represented, as shown in Figure 5.

For example,	H_2SO_4	becomes	H2SO4
	60°C	becomes	60 deg C



Note: Characters ringed should not be used in abstracts

Fig.4 The Datamatic tape-typewriter keyboard character set as used at the United Kingdom Defence Research Information Centre

1. Angstrom unit Å: replace by "A".
2. Chemical formulae: write on one line, eg H_2SO_4 becomes H2SO4.
3. Common logarithms (base 10): use "log".
4. Natural logarithms (base e): use "ln".
5. Square roots √ use "square root of".
6. &: use either "and" or "+", as appropriate.
7. ≥: use ">" or "=".
8. ≤: use "<" or "=".
9. Fractions: use an oblique stroke and brackets if necessary, eg $\frac{a-b}{c}$ becomes (a−b)/c.
10. →: use "yields" if chemistry, "approaches" if mathematics.
11. #: use "no".
12. Degrees °: use "deg".
13. Subscripts: use "sub", eg C_L becomes C sub L.
14. Exponents: use "to the power", eg x^4 becomes x to the power 4. However, numbers involving * powers of 10 may usefully be simplified, eg 1.19×10^4 becomes 11,900.
15. The ** question mark: omit in title, and rephrase abstract to make it unnecessary.
16. Greek letters: use names, in general. However, it may be more useful to name the quantity concerned, eg wavelength (λ), frequency (ν), angular velocity (ω) etc.
17. ∞: use "infinity".
18. ∫: use "the integral of".
19. %: use "per cent".
20. ½: use 1/2.

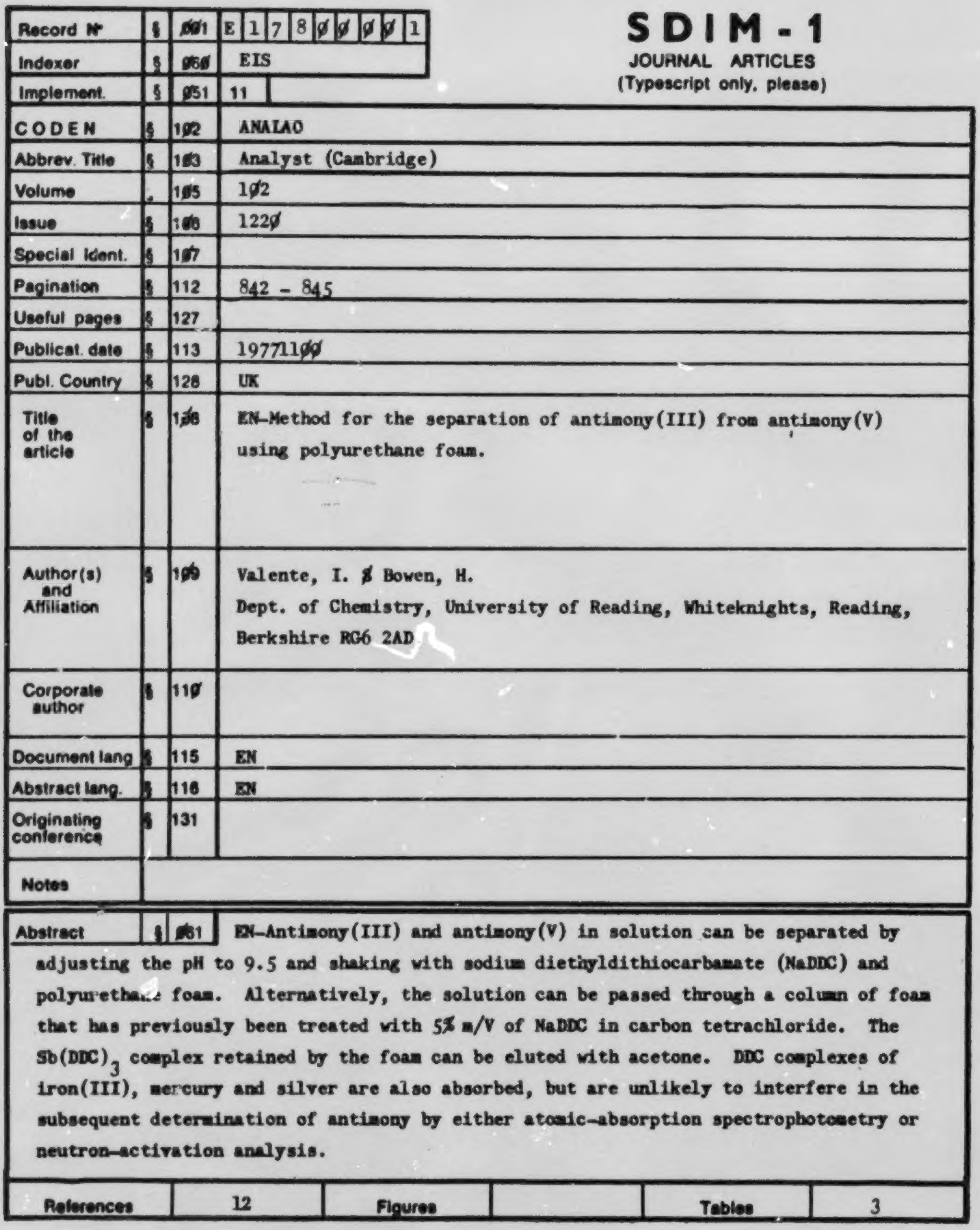
Notes: * positive or negative
** used as a field terminator

Fig.5 Representation of special symbols in the United Kingdom Defence Research Information Centre data base

3.4 Labels and Formats

A computer program which processes data needs to be able to recognise the data with which it is working. Individual pieces of information (sometimes called data elements) are usually labelled for this reason. Information is often entered onto a specially-prepared form and a place is reserved for each data element. An example of a form used for input is shown in Figure 6 (The System for Documentation in Metallurgy). Each data element is given a three-digit number (called a 'tag') e.g., CODEN has a tag of 102 in the example in Figure 6. Note that individual authors are separated by a '$' and the author affiliation appears under the same tag. Further examples are shown in Appendix B.

It is important to decide what a data element comprises. For instance, should all the authors of a document be given a single tag (as in the example in Figure 6) or should each author be given one? Increasing the depth of labelling increases the work required to input and process the information. On the other hand, if there are fewer elements, i.e. if elements which are put together only have one tag, the computer might not be able to do what the user wants because it cannot separate the elements.



SDIM - 1
JOURNAL ARTICLES
(Typescript only, please)

Record Nº	§	001	E17800001
Indexer	§	060	EIS
Implement.	§	051	11
CODEN	§	102	ANALAO
Abbrev. Title	§	103	Analyst (Cambridge)
Volume	§	105	102
Issue	§	106	1220
Special Ident.	§	107	
Pagination	§	112	842 - 845
Useful pages	§	127	
Publicat. date	§	113	19771100
Publ. Country	§	128	UK
Title of the article	§	108	EN-Method for the separation of antimony(III) from antimony(V) using polyurethane foam.
Author(s) and Affiliation	§	109	Valente, I. $ Bowen, H. Dept. of Chemistry, University of Reading, Whiteknights, Reading, Berkshire RG6 2AD
Corporate author	§	110	
Document lang	§	115	EN
Abstract lang.	§	116	EN
Originating conference	§	131	
Notes			

Abstract	§	081	EN-Antimony(III) and antimony(V) in solution can be separated by adjusting the pH to 9.5 and shaking with sodium diethyldithiocarbamate (NaDDC) and polyurethane foam. Alternatively, the solution can be passed through a column of foam that has previously been treated with 5% m/V of NaDDC in carbon tetrachloride. The $Sb(DDC)_3$ complex retained by the foam can be eluted with acetone. DDC complexes of iron(III), mercury and silver are also absorbed, but are unlikely to interfere in the subsequent determination of antimony by either atomic-absorption spectrophotometry or neutron-activation analysis.

References	12	Figures		Tables	3

Fig.6 An example of input to the System for Documentation in Metallurgy (SDIM)

The depth of labelling or tagging is also dependent on the format of each data element. Implicit labelling can sometimes be introduced into data elements by the use of standard syntax which reduces the amount of explicit tagging required. For a group of authors:

Smith, A.U; Jones, P.J.

the surname is separated from the initials by a comma, initials are separated by a full stop and authors are separated by a semicolon. The use of such standard syntax allows a program to pick out the surname of each author. Standards and guidelines for the representation of bibliographic data are available, for instance in the UNISIST Reference Manual[18] and the ISBD[19] (International Standard for Bibliographic Description). It is useful to produce a 'data dictionary' in which the content and format of each element are described, together with examples. This document will serve as a standard reference for all persons concerned. An example is shown in Appendix B.

3.5 Accuracy

Once information has been input to the computer, it must be validated and checked for accuracy (data validation). The techniques used for correction of errors will depend on the equipment available, for instance a visual display unit attached to a computer can be used to perform an immediate check. Corrections can also be made at that time unless extensive proof reading is required. Hardware aspects are dealt with in more detail in chapter four.

A number of techniques exist for using the computer for data validation. For instance, the addition of a check digit to a number allows the computer to check that a number has been keyboarded correctly. How this is done is explained in Figure 7 which shows how to calculate a check digit for an International Standard Book Number (ISBN). The technique is a valuable one because it is usually important to have an accurate and unique identification code for each item. Some data validation techniques are given in Table 1. The possible checks will depend to a large extent on the data being treated. Hines[20] suggests that an empirical approach to development of error checking routines should be adopted with emphasis on the convenience of the person doing the input rather than the programmer. Hart[21] describes the steps taken to input data for the production of an abstract journal. Details of the input, validation and error correction routines are given and it is a useful description for the organizations with little or no practical experience. The United States Defense Documentation Center[22] has produced a data input manual which describes the error routines used in conjunction with their on-line input system.

The International Standard Book Number (ISBN) is a number that identifies one title, or edition of that title if there is more than one, or volume of a multi-volume work, from one specific publisher, and is unique to that title or edition or volume. The construction of an ISBN is explained in British Standard BS 4762: 1971. The check digit of an ISBN number is used to guard against errors in transcription. It is calculated in the following manner:–

A. Each digit in the group, publisher and title number is multiplied by a "weight". The first digit is multiplied by 10, the second by 9, the third by 8, and so on.

For example,	0	8	3	5	2 0 0 0 1	– ISBN
	10	9	8	7	6 5 4 3 2	– weights
	0	72	24	35	12.0.0.0.2	

B. The products of the multiplication are added up, thus:

$$0 + 72 + 24 + 35 + 12 + 0 + 0 + 0 + 2 = 145$$

C. The result is divided by 11 to find the remainder:–

$$145 \div 11 = 13 \text{ with a remainder of } 2$$

D. The remainder is subtracted from 11 to find the required check digit*:–

$$11 - 2 = 9$$

*If the result of subtracting the remainder from 11 is 10 the check digit is written as X, thus maintaining the ISBN at a standard length of ten digits.

E. The check digit is appended to other digits to complete the ISBN, eg., ISBN 0-8352-0001-9.

Fig.7 Calculation of ISBN check digit

TABLE 1

Some Data Verification Techniques

Verification technique	*Description*
Proof reading	Manual check of the input against the original
Key data twice	Machine compares two versions (used mainly on card punch devices)
Consistency and plausibility by program:	
1. Check-digits	see ISBN example in Figure 7
2. Plausibility of data	e.g., dates in a particular range, page numbers that follow
3. Presence of certain data elements	certain elements may *have* to appear or certain combinations e.g., a document may always have to have a title
4. Syntax checks	Authors name followed by a comma, spaces after punctuation
5. Sequencing	A sequential number given to lines of data allows the sequence of lines to be checked
6. 'Hash' totals	the computer counts the number of lines or records input and compares this against a given total
7. Type of data	e.g., a data element may be all numbers or just letters
8. Field length	for fixed length data elements, that the length is given: for variable length elements, that length is not greater than maximum
9. Use of authority files	the content of data elements e.g., thesaurus terms can be checked against a master list of terms

4. INPUT EQUIPMENT

4.1 Introduction

The GIGO (Garbage in – Garbage out) principle serves only to underline the importance of having a system which ensures that information is accurately transferred into machine-readable form. The traditional methods (punched card and paper tape) are being superseded by more modern techniques including key-to-magnetic storage, on-line input and optical character recognition. This chapter describes the various equipment in use for information systems and compares them. Graphic displays are not included. The methods of input included are:

- punched card,
- punched paper tape,
- key-to-magnetic storage devices,
- on-line input,
- optical character recognition.

4.2 The Punched Card (IBM card)

The punched card was perhaps the earliest medium for recording information in machine-readable form. The technique was, in fact, used purely as a storage medium for many years before being used in conjunction with computers.

The standard punched card is divided into 80 columns and 12 rows, each column being equivalent to one character of information. Thus, one card can hold eighty characters of data and an example is shown in Figure 8. Holes are punched in the card according to the character to be recorded and the combination of holes for a particular character is

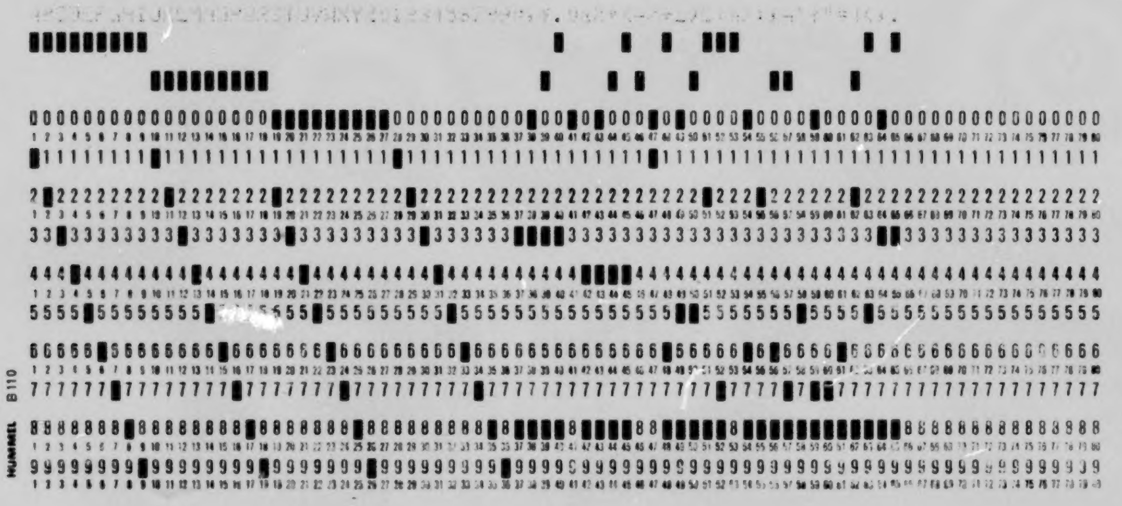


Fig.8 An example of a punched card

unique. The characters recorded are printed along the top of the card. A variation on the 80-column card is the IBM 96 column card which is much smaller in size and can record more data. However, the vast majority of cards in use retain the standard 80-column format. The machine for punching a card has a keyboard rather like that of a normal typewriter except that there are differences in the layout of the keys. The blank cards are stored in a hopper and are fed through the machine as needed.

Different machines have various degrees of sophistication and the range of facilities includes:

Verification:	an operator re-types on the keyboard the content of the cards in order to check whether the recorded data are correct. Usually a small notch is punched in the card if the data are correct.
Sorting:	the machine will sort the cards into a predefined order based on the content of one or more columns
Interpreting:	the characters corresponding to the sequence of holes are printed at the top of the card. (This is useful, for instance, if a deck of cards has been punched by a computer.)

The character set on punched card machines is usually limited in practice, for although it is possible to punch special characters, this usually has to be done by punching the combination of holes for each character individually, which takes longer. The advantages and disadvantages of this input medium are shown in Table 2.

TABLE 2

Advantages and Disadvantages of the Punched Card as an Input Medium

Advantages	*Disadvantages*
Visible check of data possible	Density of storage low
Individual cards can be sorted by machine	Can easily get out of order
Can be used as a manual reference	Relatively slow on input to computer (compared with magnetic tape)
Inexpensive medium	(Usually) limited character set
Large number of card punching machines in use	Necessary to repunch a whole card to change data

4.3 Punched Paper Tape

Like punched cards, paper tape was available before the advent of the computer and since then has seen widespread use, although it has been more popular in Europe than in the USA. Early paper tape punches had five tracks giving a total of 32 (2^5) characters only; later the number of tracks was increased and the layout of an eight-track tape is shown in Figure 9. A row of eight positions across the tape corresponds to a single character with seven positions representing the character and the eighth, a parity check. A single character is represented by a predetermined set of holes and blanks. The number of holes punched is always either odd or even, depending on the convention used, and the eighth hole is either punched or not, to produce a single parity. The technique is a checking mechanism. The small holes are called sprocket holes and are used to feed the tape through a reader, either by engaging a sprocket wheel or providing clock pulses when being read by a higher speed photoelectric reader. The relative advantages of the medium are listed in Table 3.

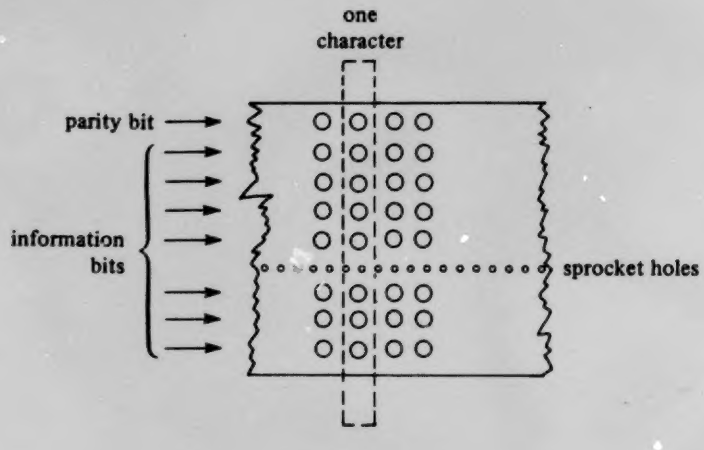


Fig.9 Punched paper tape

TABLE 3

Advantages and Disadvantages of Paper Tape as an Input Medium

Advantages	*Disadvantages*
Easy to handle	Difficult to read directly
Takes up little space	Records cannot be physically sorted
Transported easily	Need to repunch a whole record to correct error
Record always available	Relatively slow reading speeds
Variable length records	Material tends to be fragile
Inexpensive medium	

4.4 Key-to-Magnetic Storage Devices

4.4.1 Introduction

Devices which record data directly onto magnetic storage devices are a comparatively recent commercial innovation. There is now a wide variety of machines available. Magnetic tapes, floppy discs and discs are all used as storage media. (Magnetic storage media are described in chapter five). The machines often have a processing capability, allowing data to be validated before being stored. In effect they are a small computer. Some systems use a hardcopy device (like a typewriter) for displaying input, others use visual display units. It is possible for a number of terminals to share a single processor.

An outline of a typical system is shown in Figure 10 which shows the inside of a DATAPOINT* key-to-cassette tape processor. The system has a typewriter keyboard plus a separate numeric pad of 11 keys. A video display of 80 columns by 12 rows (or alternatively 80 by 24), giving 960 (or 1920) character positions, is provided. Information is stored on cassette tapes and in some cases there are two cassette tape recorders in a single machine.

---

* DATAPOINT is a copyright held by the Datapoint Corporation.

Inside the DATAPOINT processors

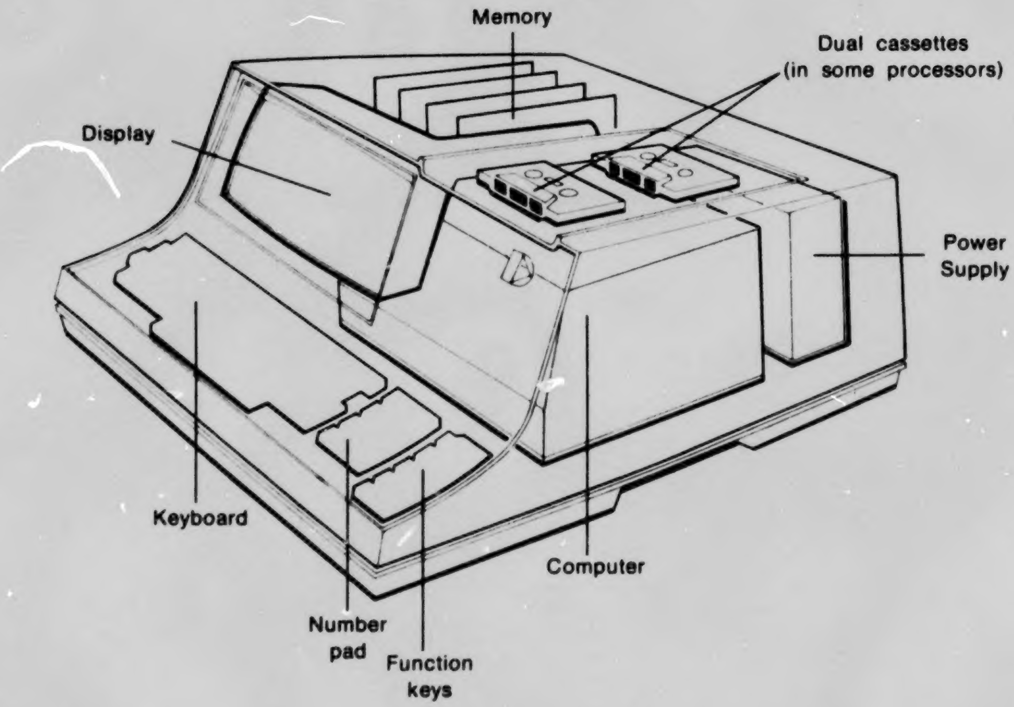


Fig.10 Schematic diagram of a DATAPOINT® key-to-cassette tape processor
(from the Datapoint Equipment Catalogue 1978 Copyright © Datapoint Corporation)

4.4.2 Hardware

There is a wide variety of hardware to choose from, ranging from a single station (see Figure 11) with one keyboard, typewriter output and simple cassette store, designed to replace punched card and tape, to powerful multistation devices with several keyboards attached to the same processor. The following shows the range of possibilities:

Processors/Main Memory

From a small buffer (for storing the characters displayed on a screen) upwards – the larger the processor, the more powerful the machine becomes.

Keyboards and character sets

Either traditional keyboard or typewriter layouts – character sets can be upper and lower case and special characters can sometimes be added.

Storage devices

Cassette tapes hold approximately 100,000 characters. Diskettes (floppy discs) usually hold 250,000 characters upwards, depending on packing density and the number of read-heads. Disk units, with a wide range of storage capacities are also available.

Display units

Typewriter output or visual display units of varying sizes are usual.

Printers

A range of printers is available with extended character sets, with speeds from tens of characters per second to hundreds of lines per minute. It is also possible to program for certain special characters with the matrix printers available.

Communications

Communications facilities, linking devices to a computer are available so that devices can act as terminals or send batches of data (remote job entry) to a central computer.

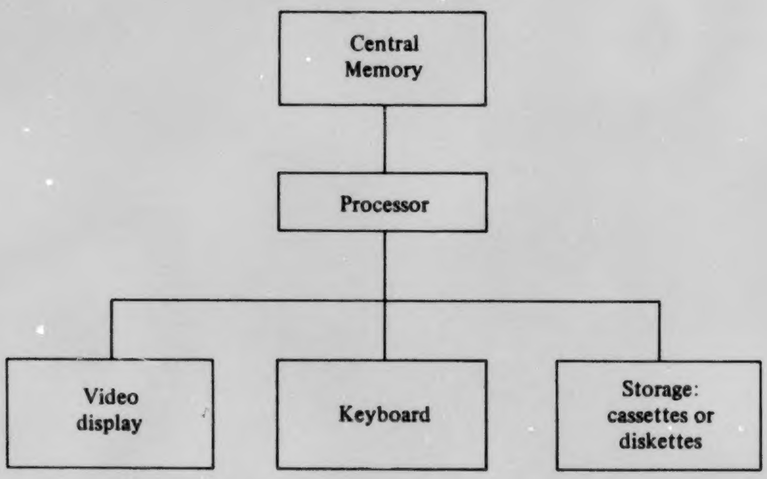


Fig.11 Components of a basic key-to-storage system

4.4.3 Capabilities of Key-to-Storage Devices

Data are entered using the keyboard and are displayed on the screen, if one is available. If the operator sees an error it is usually possible to correct it immediately. With the VDU the screen can be arranged in such a way that data element names are already presented to the operator. In some systems, label or tag information is held in protected zones on the screen which can not be overwritten by the operator. A further advantage of this technique is that only the data needs to be entered and not the tag information. Once the data are validated it is possible to correct them immediately on the screen. Following this step, they can be stored.

Communication with a central computer can be done either by having a direct communications link between the key-to-storage device and the computer or by using a compatible medium. For instance, data are often transferred to half-inch magnetic tapes using a converter. The possibilities depend on the equipment available, e.g., the Nixdorf 620/15 and 620/45 (shown in Figure 12) transfers the information directly to half-inch tape.

The larger the machine, the greater is the range of operations that can be performed. Figure 13 shows the elements of a multi-work station shared processor system. Larger processors offer more possibilities for validation and record manipulation. For instance, it is possible to have stored programs and use them to operate on the data in real-time. If compilers are available, tailor-made programs can be written for particular applications. Shared processors offer other advantages over single units. For instance, certain equipment (a printer, for example) need only be obtained once. However, if shared units fail, the whole system can be put out of operation and hence the level of duplication of equipment (e.g., discs) needs to be thought out carefully.

4.4.4 Comparison of Key-to-Storage Devices with Punched Tape and Cards

Key-to-storage devices were developed to replace the punched card and paper tape input systems. Although they are more expensive, they do provide a higher throughput of material and in many cases an increased 'operator morale'. Data can be entered much more quickly than with the card punch because of the error correction and verification facilities and because the number of characters to be entered can be smaller as, for instance, tag information can be pre-set on the screen. Major disadvantages of key-to-storage devices seem to be the cost, if a more sophisticated system with printer and VDU are acquired, and the lack of hardcopy on the simpler devices. The skill of the operator needs to be greater too. A summary of the advantages and disadvantages appears in Table 4.

4.5 On-Line Input

4.5.1 General

On-line input can be defined as the input of data directly to the machine in which they are stored. Usual equipment for input are the typewriter type terminal and visual display units (VDU). The central processor is used for data verification and editing. There is a wide variety of terminals available: Axner and Reaghan[23] indicate some 170 products from 76 vendors on the USA market.

The major advantage of the technique compared with key-to-storage systems is the fact that it is possible for the operator to interact directly with the information files stored in the mainframe computer. For instance, it is possible to check indexing terms against a thesaurus immediately the data are input. Error corrections can be carried out quickly.

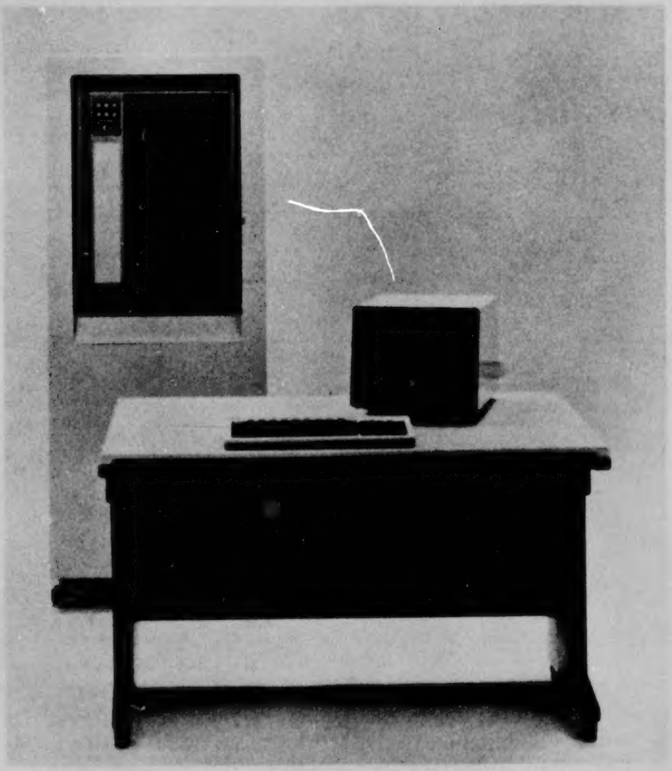


Fig.12 The NIXDORF key-to-storage input equipment (model 620 with half-inch magnetic tape)

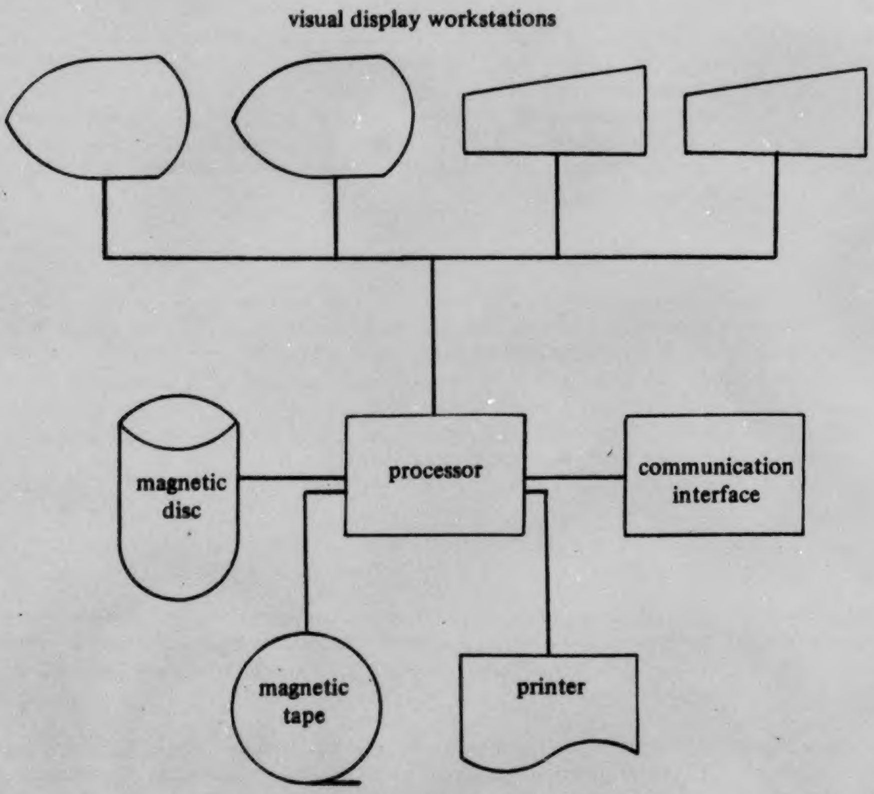

visual display workstations
magnetic disc
processor
communication interface
magnetic tape
printer

Fig.13 Elements of a shared processor key-to-disc system

TABLE 4

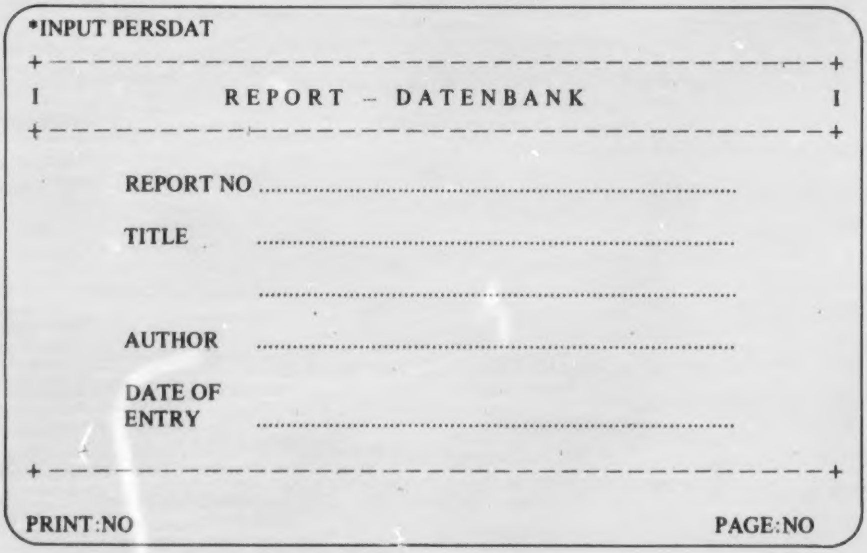
Some Advantages and Disadvantages of Key-to-Storage Devices

Advantages	*Disadvantages*
Higher throughput c.f. punched card	Relatively higher equipment cost but shared processors reduce cost
Immediate validation and error correction	No immediate hardcopy on simpler devices with VDUs
Formatted screen facilities – reduction in number of characters to key	Shared processor facilities can mean that all operators are idle
Data quickly available for use	Greater operator skill required c.f. punched card
Low noise level	
Easier handling of physical files	

On the other hand, it would be necessary to bear in mind the costs of such an input method: for instance, the overhead for transaction processing in the computer and having the files on-line for long periods. Key-to-storage devices can also be on-line, allowing data to be entered locally and partly verified; later the verification can be completed in the mainframe computer.

4.5.2 Software

Some manufacturers produce general purpose software for input. If not, or if special facilities are to be provided e.g. verification of the correctness of indexing terms, software will have to be specially written. The IDRC[10] minicomputer-based bibliographic information system provides such facilities. Appendix A shows an example of the dialogue between the computer and the operator. Another example of an on-line input system can be found in the GRIPS[9] system. The facilities of this software package for the Siemens 7000 series are shown in Figure 2. The central module is a data base management system for the maintenance of data bases. The input module (DINUPS – DIMDI Input and Updating Software) allows the creation of a pre-formatted record grid for various applications. An example is shown in Figure 14. Data can be entered into the data base and edited using DINUPS. The advantage of the on-line input comes when the retrieval software DIRS2 (DIMDI Information Retrieval System – Version 2) is used. It is possible to use the DIRS2 software immediately data are entered to check that they are correct. Thus the system allows a full interaction between the operator entering the data and the retrieval software by first using DINUPS and then DIRS2.



```
*INPUT PERSDAT
+------------------------------------------+
I           REPORT - DATENBANK             I
+------------------------------------------+

        REPORT NO  ........................

        TITLE      ........................

                   ........................

        AUTHOR     ........................

        DATE OF
        ENTRY      ........................

+------------------------------------------+

PRINT:NO                            PAGE:NO
```

Fig.14 An example of a pre-formatted VDU grid for the DIMDI DINUPS input system

4.6 Optical Character Recognition

4.6.1 Description

Optical character recognition (OCR) is a method of data entry which converts data into machine-readable form without the need to use a keyboard. Generally, OCR readers are divided into the following three categories:

- Optical character readers (OCR)
  Machines have varying levels of sophistication but the best ones can read an extended character set of letters and numbers which can be typed, printed, handwritten or output from a computer.
- Optical bar code readers
  These machines read a series of lines which represent a number.
- Optical mark readers
  These machines read marks placed in predefined areas of a page.

Optical character readers are clearly the most sophisticated of the three categories of machine. Optical bar code readers can be used, for instance for library circulation systems, the accession number and reader number being represented by a series of lines, the former in the book and the latter on an identity card. Optical mark readers recognise the presence or absence of a mark at a predefined point on a page. They are frequently used for survey work, questionnaire processing etc. Auerbach[6] contains a useful review of OCR equipment.

4.6.2 Character Sets

Optical character recognition equipment can be made to read typescript, printed pages, handwritten material, computer output and computer output microfilm. The typescript usually has to be in one or more specified formats. Those recommended by the International Organization for Standardization are called OCR-A and OCR-B[24,25]. The character set for the latter is shown in Figure 15. Such fonts can be obtained on exchangeable writing heads, for instance as used on the IBM typewriter.

A number of other fonts are in use such as the IBM 1403 which is that used on the 1403 line printer. 3/16" Gothic is also a font used for optical character recognition. At INSPEC (The Institution of Electrical Engineers) a COMPUSCAN OCR equipment is used with the standard COMPUSCAN character set. An example is shown in Figure 16. Some machines such as the IBM 3886 are capable of reading handwritten material.

4.6.3 Make-up of an OCR Reader

An OCR reader basically comprises:

- an input mechanism,
- a scanning and recognition mechanism,
- an output mechanism (for the processed documents),
- a controlling mechanism,
- a link to a data recording device or directly to a computer.

An example is shown in Figure 17. The input mechanism depends on the form of the input which can be:

- journal tapes and strips (for example, a cash register),
- document readers which read lines or sections on a page,
- page readers which read multiple lines of type on a page,
- hand-held wand readers.

The scanner and recognition subsystems read the characters and convert them into electrical signals which are then analysed.

The Control Data Corporation model 92650 is an example of an OCR page reader. The macine can adapt to a variety of applications such as input to an intelligent terminal, remote input device to a central computer, multi-media (OCR key-to-disc system), standalone OCR data entry system and input device to a communications system. OCR-A and OCR-B fonts are available. Instantaneous read rate is 683 characters per second throughput or one full page every 12 seconds. The Siemens HL5 hand-held OCR scanner is an example of a portable OCR scanner which might be used for a library circulation system. It recognises a number of standard alphanumeric fonts.

4.6.4 Feasibility of the OCR Solution

OCR clearly has an advantage over other methods of input when original forms can be used, for in this case it is not necessary to re-type the data. Hence, OCR has clear advantages for documents which are produced by the computer (such as bills or orders) and which have to be further processed after data have been added.

Y.4.2 Illustration of OCR-B characters (complete set according to ISO Recommendation R 1073) scale 5:1.

Y.4.2.1 Characters of OCR-B having a strokewidth of 0.35 mm (0.014 in) for sizes I and II and of 0.38 mm (0.015 in) for size III.

ABCDEFGH
IJKLMNOP
QRSTUVWX
YZ*+,-./
01234567
89
ÄÖÅÑÜÆØ
↑≤≥×÷°

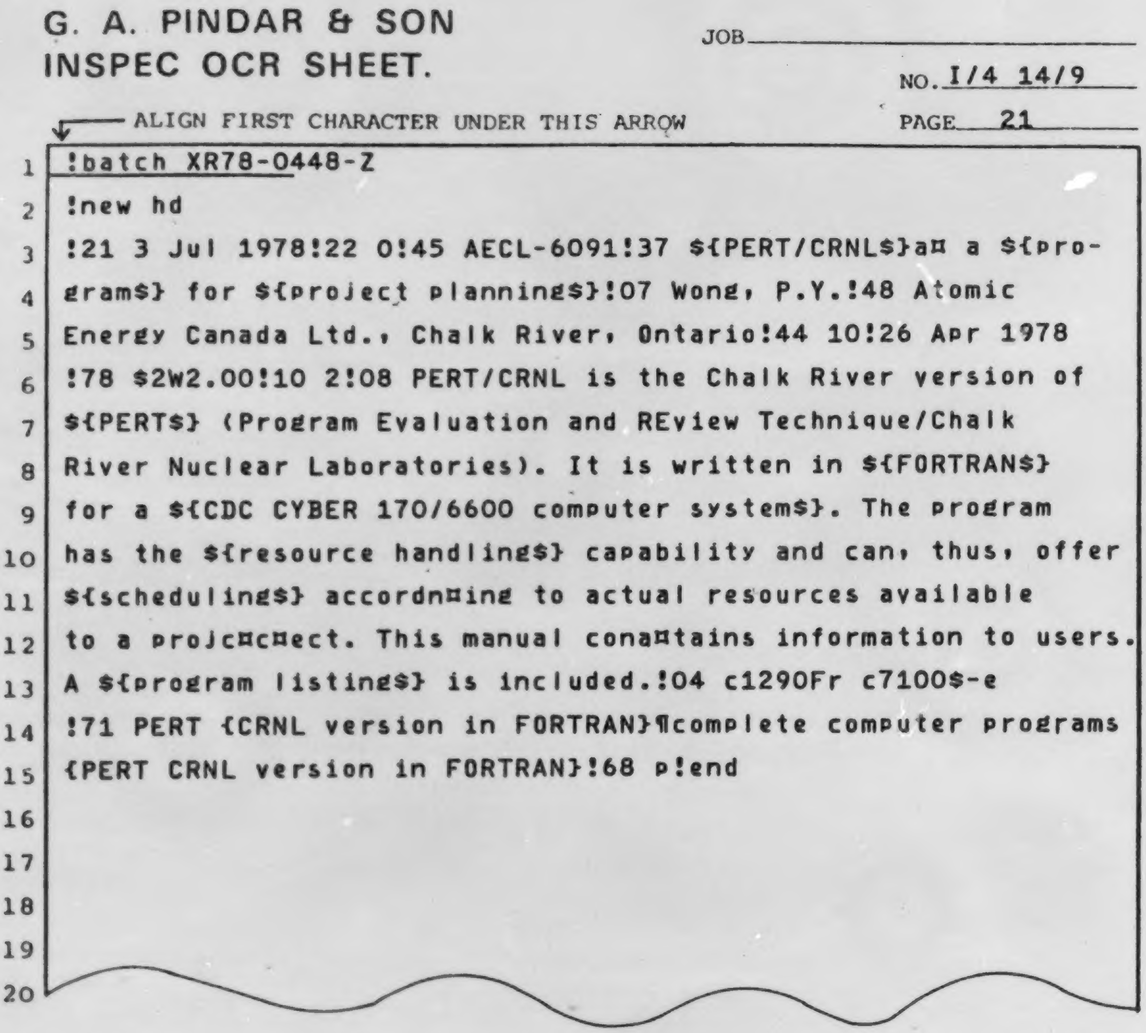
Y.4.2.2 Character of OCR-B having a strokewidth of 0.31 mm (0.012 in) for sizes I and II and of 0.34 mm (0.013 in) for size III, which are considered in clause 4.5.

£$:;<%>?[!#&¤]
(=) _

Y.4.2.3 Characters of OCR-B having a strokewidth of 0.31 mm (0.012 in) for sizes I and II and of 0.34 mm (0.013 in) for size III, which are considered in clause Y.2.9.

abcdefgh
ijklmnop
qrstuvwx
yz m åøæ
¨ ´ ` ^ ~ ˇ
@ '

Fig.15 The OCR-B character set of the International Organization for Standardization (ISO)



G. A. PINDAR & SON
INSPEC OCR SHEET.

JOB

NO. I/4 14/9

ALIGN FIRST CHARACTER UNDER THIS ARROW

PAGE 21

```
!batch XR78-0448-Z
!new hd
!21 3 Jul 1978!22 0!45 AECL-6091!37 ${PERT/CRNL$}a¤ a ${pro-
gram$} for ${project planning$}!07 Wong, P.Y.!48 Atomic
Energy Canada Ltd., Chalk River, Ontario!44 10!26 Apr 1978
!78 $2W2.00!10 2!08 PERT/CRNL is the Chalk River version of
${PERT$} (Program Evaluation and REview Technique/Chalk
River Nuclear Laboratories). It is written in ${FORTRAN$}
for a ${CDC CYBER 170/6600 computer system$}. The program
has the ${resource handling$} capability and can, thus, offer
${scheduling$} accordn¤ing to actual resources available
to a projc¤c¤ect. This manual cona¤tains information to users.
A ${program listing$} is included.!04 c1290Fr c7100$-e
!71 PERT {CRNL version in FORTRAN}¶complete computer programs
{PERT CRNL version in FORTRAN}!68 p!end
```

Fig.16 Input of bibliographic information to the COMPUSCAN OCR reader (UK – INSPEC)

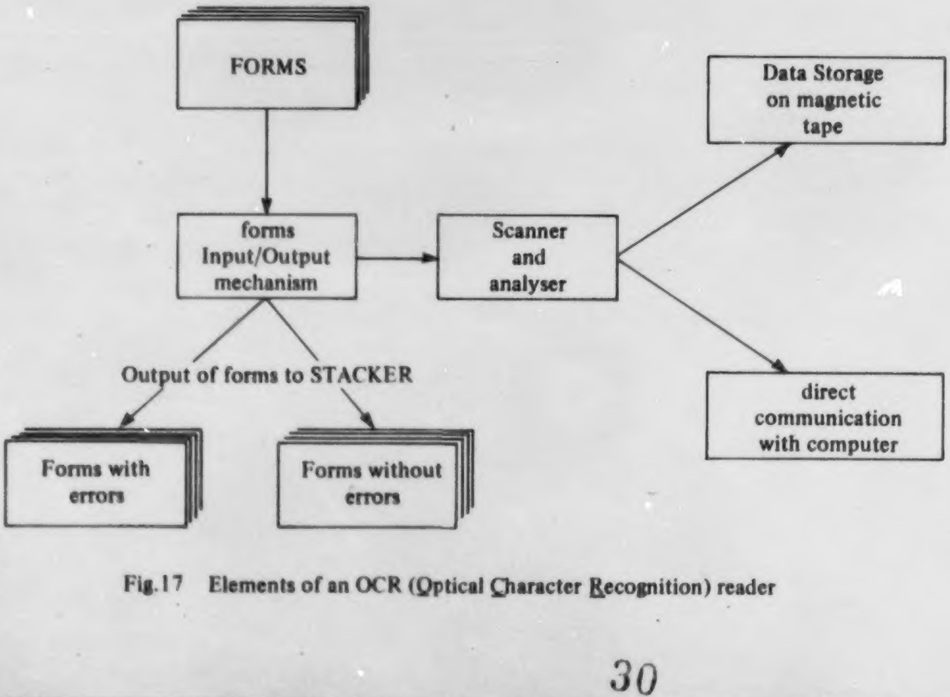

FORMS
Data Storage on magnetic tape
forms Input/Output mechanism
Scanner and analyser
Output of forms to STACKER
direct communication with computer
Forms with errors
Forms without errors

Fig.17 Elements of an OCR (Optical Character Recognition) reader

As a straightforward replacement to the keypunch or key-to-magnetic storage device, the advantages of OCR are less clear, if a data entry form has to be specially created. The relatively high cost of OCR equipment means that a high volume of input is necessary to justify the machine unless it is shared with other users.

In choosing an OCR machine, the needs of the applications must be balanced against the level of sophistication offered by the equipment. Benchmark tests can be devised to compare a machine with its rivals. Factors to look out for include:

- error rates and reject rates,
- throughput,
- number of different fonts read,
- input formats (page, document, journal).

The use of normal typewriters is a clear advantage over other input equipment for operators do not have to be specially trained. OCR equipment is sometimes linked with key-to-disc equipment or a visual display terminal linked to a mainframe computer. This allows data to be validated and edited on-line.

4.7 Choosing Input Equipment

Often it is necessary for budgetary reasons to make use of equipment already available in an organisation (for example, a card punch shared with other applications, an existing visual display terminal or a central data-processing department). For those who have to choose new equipment, however, there are a number of criteria which will limit the choice and some of these, are given in Table 5. General advice on project management is given in Chapter six. The first step to undertake is the definition of the needs. These can be obtained from a functional description of each system for which data are needed. This can then be used to develop a functional specification of the input needs, e.g., volume of data, speed of updating required, etc.

TABLE 5

Steps to be Taken in the Choice of Input Equipment

1. Define what the needs are:
    - details of input systems forms, data elements etc. (undertake a functional analysis)
    - specify the character set
    - constraints (e.g. computer to be used, communications facilities and peripherals available)
    - volume of data
    - time constraints
    - geographical constraints

2. Examine available equipment which would fulfil needs:
    - reputation of manufacturer
    - number of installations (in your area)
    - size of company
    - delivery dates
    - documentation
    - user support (maintenance)
    - price and method of payment (purchase, lease. .)
    - use of a service bureau

3. Note details of each system:
    - hardware specification
    - software facilities
    - upgrading or downgrading options – ease of transfer from one to the other
    - character set
    - is output compatible with the mainframe computer?
    - size of machine

4. Look at cost/performance estimates for each system

5. Choose machine

An examination of the equipment available can then be made, especially a comparison between the facilities offered and those required. Information on manufacturers' products can be obtained from the manufacturer, vendors, equipment reviews, etc. Auerbach Computer Technology Reports[6] contain descriptions of available products and discuss selection criteria. There are a number of selection criteria which will depend on the weight attached to each point. For instance, it is pointless installing a sophisticated machine if the nearest service agent could not come for one week following a break down. Likewise, a reputable company which is likely to stay in business is important.

The reputation of the manufacturer is a good guide and the number of active installations is usually proof of success. It is always possible to contact other users and ask them their opinion of the equipment; a good manufacturer will provide a list. User support for training courses and documentation is important, especially if the equipment is more sophisticated. The use of a service bureau will also need to be investigated especially for OCR machines.

The method of payment will need to be considered carefully. Outright payment may be cheaper but maintenance might not be so good, there is less of an incentive to keep the machine running than if leasing or rental are opted for. Short leases give more flexibility and the opportunity to change to a more modern machine. Longer leases are usually cheaper per month. Different machines can be compared for cost against performance. Staff morale is another factor, key-to-disc machines may give more job satisfaction than the punched card machine, but costs may be higher.

5. INFORMATION STORAGE DEVICES

5.1 Introduction

Computer systems have a number of storage devices which have differing levels of performance, cost and storage techniques. Storage devices are characterized by:

- the cost per byte of storage,
- the size of the store,
- the access time, the time to retrieve the data,
- the data transfer rate.

The major types of memory in use are:

main memory, magnetic drums, magnetic discs, magnetic tapes, mass storage devices, holographic and photographic memory.

The different storage media, because of the wide range in cost, capacity and performance, are used for different purposes.

5.2 Main Memory

Main memory usually has an access time of less than a microsecond, the speed depending on the technology used. The most well-known is the ferrite core which is a core of magnetic material which can be magnetized in one of two directions. There are other ways of employing magnetic material, such as thin films and plated wires. Magnetic storage is rapidly being overtaken by semiconductor storage. Using the techniques of large-scale integration (LSI) and very-large-scale integration (VLSI) a large number of memory circuits can be built into a single 'chip' of silicon. Main memory sizes up to 64K bits on a single chip are now being produced and used in several computer systems.

The physical size of main memory is usually up to several million bytes (megabyte) with a purchase cost of several tens of thousands of dollars per megabyte; it is used for the storage of programs which are currently being executed and for data undergoing processing. The operating system is also stored here. Many operating systems can execute a number of programs at once, using techniques called multiprogramming and time-sharing. These programs are stored in different parts of the main memory. Size of the main memory plays a large part in determining the throughput of the computer e.g. because more programs can be executed concurrently and larger and more sophisticated operating systems can be employed. A number of independent manufacturers are producing 'add-on' memories which can be added to existing machines. Sometimes prices compare favourably with those of the computer manufacturer.

5.3 Direct-access Devices

Magnetic drums and discs are called direct access devices because of the ability to access data directly rather than sequentially as when using a magnetic tape. In fact the speed of access of data is significantly slower than for main memory.

The fastest of these devices is the *magnetic drum* (see Figure 18) which has a cylindrical shape and rotates at several thousand revolutions per minute. The cylinder is coated with iron oxide and a series of heads which write and read information, are placed along the drum. These heads do not move and hence these devices are sometimes called 'head-per-track' devices because data are recorded on a band ('track') of the drum covered by a single head. Access time is normally a few milliseconds and transfer rates are greater than 1 million bytes per second. Storage capacity is typically in the tens of megabytes range.

Many computers use 'virtual memory' systems which in effect expand the size of the main memory. In practice, main memory is split up into sections, often called pages and these are moved from main memory to drum storage when

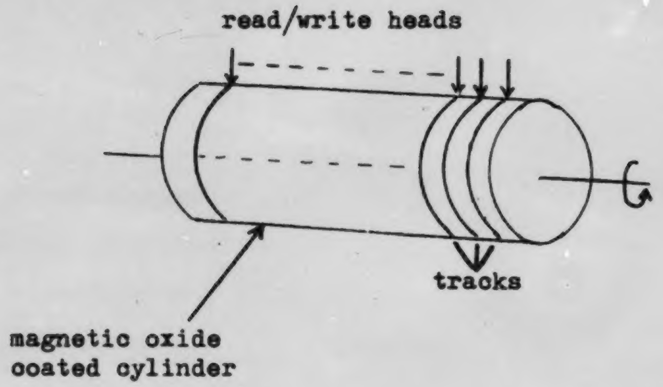


Fig.18 Outline of a magnetic drum

they are not immediately required. Drums are also used for files of information which are in frequent use, the keeping of which in main memory is not justified.

Magnetic disc devices consist of a series of flat circular platters which are held horizontally, are stacked one above the next and which are rotated about a common vertical axis. Each platter is coated with magnetic material. There are two methods of reading and writing, with fixed or with moveable heads. In the fixed head device the heads are arranged as shown in Figure 19. As with the drum, there is one head per track and hence the performance of these devices is similar. Moving head discs have one read head per platter side (see Figure 20). The read heads are attached to a series of arms which are connected together to form a comb. To read or write to a particular track, the comb has first to be positioned on the appropriate part of the disc. All the read/write heads move together so data are usually written in 'cylinders' ie. all the tracks which can be written without moving the heads.

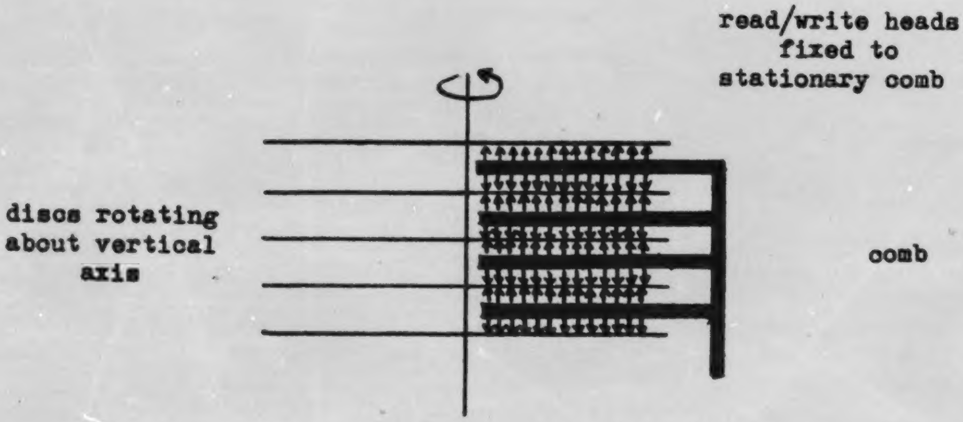


Fig.19 Section through a fixed-head magnetic disc storage unit

Access time for discs is slower than for drums; this time is dependent on the seek time (the time to find a particular track) and the rotational delay. Some discs can be removed from the drive, increasing the storage capacity of the machine. Capacities vary from several megabytes (single platter cartridge) up to several hundred megabytes (multiple head, multiple platter). Average access times of around 20 msec, a transfer rate of 1.2 megabytes per second and a capacity of 635 million bytes are available.

Even smaller capacities are available on the flexible discs, sometimes called floppy discs or diskettes. The flexible disc is a circular piece of polyester film of about 5 inches (or larger) which is coated in magnetic iron oxide and covered in an outer plastic or paper jacket. Capacities depend on recording density and whether both sides of the disc are utilised, but over 1 megabyte can be stored on a single disc.

Discs play an important part in the operation of time sharing and multiprocessing systems by providing fast access to larger amounts of data. They are used in library and information centres for storing large files of information. For

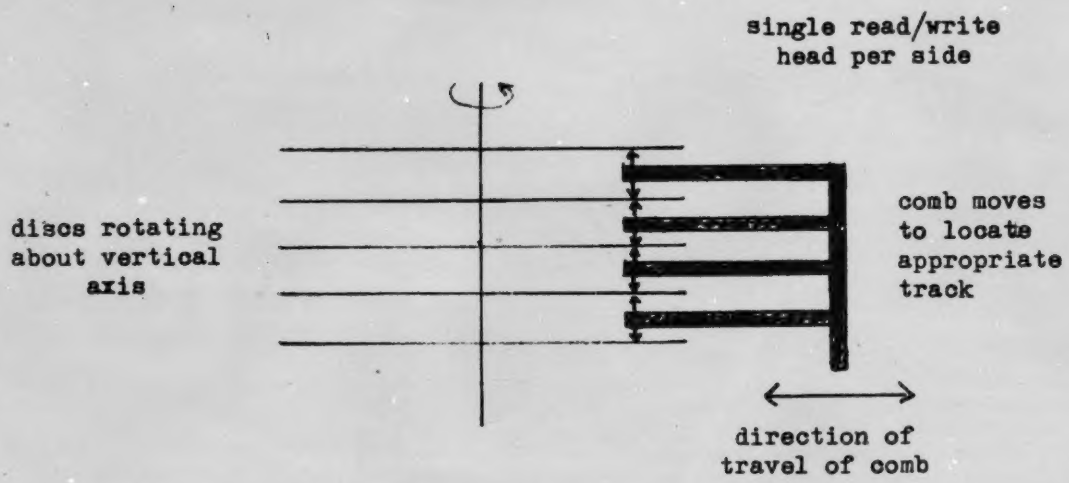

single read/write head per side
discs rotating about vertical axis
comb moves to locate appropriate track
direction of travel of comb

Fig.20 Section through a moving-head magnetic disc storage unit

instance, millions of bibliographic references are available on-line from data base services. Because of the ability to retrieve information quickly, a customer can for instance, obtain bibliographic details of articles of interest within a few seconds. Costs are kept competitive because many users can be served simultaneously. The disc storage facilities available on mini-computers are also improving and are now extending the possibilities of on-line operation for smaller organizations.

5.4 Magnetic Tape

Magnetic tape is a recording medium known to most people because of its use for recording music. In the computer world the ½ inch 9-track tape has become standard. The normal 2400 ft tape can store over 40 million bytes when recorded at a density of 1600 bytes per inch. Tapes of shorter length are available, as are cassette tapes which are frequently used for input of data. The high capacity, low cost and reasonably high data transfer rates make magnetic tape a valuable medium for archival storage and for storing files of information which have to be accessed sequentially. However, access time is very high because of the time taken to mount the tape in the first place and because the tape has to be unwound to reach the required part.

5.5 Mass Storage Devices

The mass storage device provides an answer to the problem of handling magnetic tapes. The equipment consists of an array of magnetic cartridges housed in a honeycomb shaped store. The capacity of the Control Data Corporation cartridge is 8 million bytes. When a particular cartridge is requested, an electromechanical device retrieves the requested cartridge and transfers the data to a disc. Storage capacities of the IBM 3850 and CDC 98500 systems are high; IBM claim between 35 billion and 472 billion bytes and the average access time is seconds. CDC claim a 3.8 second average access time.

A somewhat similar, but now largely outmoded device is the data cell which retrieves the selected strip of flexible magnetized material and fastens it to a rotating drum.

5.6 Photocopy and Laser-Holographic Storage

Large-capacity, direct-access storage devices exist which use laser technology.[26] Information is written onto photographic film or a holographic substrate using a laser. Hence, the store comprises a magazine of holographs or films which are retrieved in a fashion similar to that of the data cell.

5.7 Newer Memory Types

Other types of storage devices are coming onto the market. A charge-coupled device (CCD) was recently announced[27] and is intended to be a replacement for magnetic drums. CCD and bubble memories are designed to fill the gap between main memory and direct access storage, providing access to data in the millisecond to microsecond range.

Storage is likely to decrease in cost and increase in performance in the next few years. Discs with over 1000 megabytes capacity are likely to be available with transfer rate of several megabytes per second and access time less than 20 milliseconds. Costs of storage may well be less than $500/megabytes/month for main memory, bubble or similar less than $50, discs $1, mass storage less than one cent (USA Currency).

6. THE MANAGEMENT OF PROJECTS

6.1 The Project Environment

The introduction of a new system, be it an input or storage system described above or another type of system, usually involves changes within the organization. The introduction of change can be difficult and costly unless the change is carefully planned and executed. Many organizations introduce new systems, using the concept of a 'project'. A project can be defined as a temporary organization of resources to achieve a particular goal. The project usually has a leader or manager whose task it is to see that the project is carried out successfully and within the budget.

An important aspect of a project is the project plan, which will usually be devised by the project manager. Projects are often split up into phases (see paragraph 6.2) and at the end of one phase, a detailed plan for the succeeding phase is drawn up. An overall plan, produced at the beginning of the project, should be updated at the end of each phase. Plans should include details of:

- project costs (including cost/benefits analysis),
- timescale (schedules for completion of project, milestones, etc.,),
- manpower requirements,
- equipment requirements.

A specification of the work to be done should be drawn up.

The planning documents together with the results of a preceeding step, allow a decision to be made on whether to go on to the next phase of the project. If any decisions on alternative paths have to be made, management is able to make a decision with the maximum information available, including detailed information on costs. In any case, the decision to go ahead is only to the end of the next step.

6.2 The Systems Approach

6.2.1 The Stepwise Approach to System Implementation

It is now recognized that new systems can be efficiently introduced or old ones changed using a stepwise approach which maximizes the chances of implementing the correct system, allows management to make informed policy decisions and minimizes the effort spent if changes are made or if the project is cancelled. For information system projects it is usual to consider the following consecutive steps:

Conceptual stage:	the project is considered for adoption
Initial stage:	a minimum of effort is expended to provide enough information to decide whether it is worthwhile proceding with the project
Systems analysis and definition:	the customer and the organization carrying out the project agree on what the system should do
Design:	the system is designed i.e. a detailed description of how the system is to be constructed is drawn up
Construction and installation:	the system is built and installed
Evaluation:	a periodic review to see whether the system is performing according to specification and desires

Three of these steps i.e. systems analysis and definition, design, and construction and installation are briefly described below. For specific details, the reader is advised to turn to standard texts (see list of references)

6.2.2 Systems Analysis and Definition

The systems analysis and definition phase of a project is perhaps the most important of all the steps outlined above. It is at this stage that the needs are specified in detail. It therefore offers an opportunity for management to look at the goals of the organization. With the introduction of a new system, it is possible to either carry out the present services more efficiently or introduce improved services instead. For instance, a computer-based system could offer an individualized selective dissemination of information service (SDI), if this were not already being offered, in addition to production of a current-awareness bulletin.

The tasks required to be carried out at this stage of the project include:

- an examination of objectives of the existing and the proposed system,
- the degree to which current objectives are being met,
- a revision of objectives if necessary,
- an examination of the needs of the customer (department which is paying for the system) and the the user,
- the preparation of a functional specification which describes what the new system should do.

The functional specification or system definition serves as a form of agreement between the organization carrying out the work and the customer. It should include:

- a general description of the system including objectives and functions,
- a cost benefit discussion,
- details of information flow and data: input data (including source and method of creation), processing needs, output requirements (including distribution),
- audit controls,
- time constraints,
- 'people' factors (e.g., definition of responsibilities),
- contents of files and updating details.

A plan for the design phase should be drawn up and the overall project plan updated.

6.2.3 Systems Design

The main objective of the design phase is to produce a detailed description of the system to be implemented. Individual details of the system have to be specified, e.g., program specifications, files, details of output and input forms, hardware and special software to be used and information flow. An important part of this stage is the system flowchart which describes the flow of data through the system. Some organizations are now using the HIPO (Hierarchical Input Processing Output) technique[30] instead of a flowchart. To back up the flowchart, a description of each operation is drawn up.

The work to be done at this stage will depend on the strategy adopted – for example, whether software can be acquired or programming needs to be done, whether hardware is already available, and the complexity of the system. Grosch[28] considers that before any software or equipment is selected, a broad definition of the application and its environment should be made, including:

- explanations of tasks to be performed,
- definition of the specific data to be used,
- response requirements of the system,
- the size of the data base and growth rate,
- the number and geographical location of terminals.

The availability of suitable software is more and more determining the choice of machine. Software has an even greater impact if a dedicated computer is to be acquired. Some of the factors to be taken into account in choosing software are discussed in Chapter two.

Points to be considered in the design include:

- details of output required (e.g., formats, quantity, distribution),
- details of data elements,
- processing and logical operations (system logic),
- audit and control factors,
- details of input (e.g., formats, sources of data),
- details of files, storage devices, updating factors, data volume, response times,
- specification of computer programs,
- specification of manual procedures.

It is important to consider the human interfaces, especially the problems of system operation and user interaction.

The design specification should be included in a report which will also contain a plan for the following stage i.e. implementation. The master plan for the project should also be brought up to date.

6.2.4 Construction and Installation

During this phase the system is built and commissioned. The documents used to effect this implementation are the results of the previous phases, especially the design specification and the functional specification. Tasks to be performed include:

- writing computer programs,
- file creation,
- documentation,
- acquisition of equipment,
- training in the use and operation of the system,
- phase in, comprehensive testing and proving the system.

It is likely that this stage will be the most difficult from the planning point of view. Hence the plan produced at the end of the design stage should be detailed and at the same time flexible. The implementation period is also critical for determining whether the system meets the requirements. Comprehensive and thorough testing, operating the new system in parallel with the old, user training and systems evaluation techniques are some of the activities required at this stage in order to determine performance and to ensure satisfactory operation of the new system.

6.3 Personnel

The personnel needed for the introduction of a computer application are usually specialists in a particular area (e.g. systems analysts, programmers). If the necessary people are available within the organization, this represents the best solution because they are likely to understand the particular needs of the organization and the expertise gained will not be lost at the end of the project. If not, there are other possibilities:

- staff can be hired from consultancy organizations and software houses,
- the whole or parts of the project could be contracted out to another organization.

The major difference between the two approaches is that in the former, manpower can be hired and fitted into a project team. This is likely to be the best approach if project managers with the correct background are available in-house, because staff can then keep control of the project and can learn from the temporary hired personnel.

It will be necessary to decide on the degree of involvement the in-house staff will have after the system is installed, for instance, who will maintain the system. The type of personnel likely to be needed for systems development are:

Project Leader:	Management capability, understanding of application and of computer systems
Systems analyst:	Experience of techniques of systems analysis plus understanding of the application and the computer systems, knowledge of existing similar applications
System designers:	knowledge of software, programming techniques, hardware, communications
Programmers:	knowledge of programming, data base management systems (if used)
Computer operators:	ability to operate the system
Data base manager:	probably a librarian or information specialist with knowledge of the computer and data base management systems who is in charge of the data bases in the computer

The people who will operate and use the system will also be involved in the systems development. It is sometimes possible to obtain people who can undertake two or more of the above functions, for instance, systems designers are sometimes used for programming. This is a developing trend as data base software becomes more specialized.

It is important that the computer system, once operational, is not dependent on one person. How often are programs understood only by the person who wrote them? It is not a good practice to let this happen. The distribution of tasks among a group of people and the provision of good documentation are essential and in any case lead to better job satisfaction and staff flexibility.

6.4 Other Project Management Considerations

The above treatment has been necessarily brief and there are a number of salient points in addition to those discussed above, which should be borne in mind when embarking on an information processing project. Some of these points are listed below, but the subject is covered in greater depth in relevant works cited in the list of references, especially by Gildersleeve[31], Cleland[32] and Bingham[33]. In particular:

- the project environment should be encouraged;
- the project leader should be given the necessary power and support to carry out a project and the job should carry the requisite rewards;
- the customer (the department requesting the work done) should pay for the system;
- the individual chronological steps of the project (outlined in paragraph 6.2) should be kept separate and cutting corners should be avoided;
- planning documents should be produced before considering the next step;
- a regular reporting procedure should be introduced;
- the project leader should keep all interested parties informed;
- the specification of what is required should be frozen and agreed to in writing;
- high level management should be involved in the analysis;
- a standard for documentation and methods of work should be established and maintained (e.g. structured programming techniques);
- check points (dates) for regular reviews of the project should be introduced;
- project members should be motivated by assigning them tasks which are not too difficult, but which will develop them, giving them as much independence as possible and producing written appraisals of their performance;

- panic situations should be avoided;
- project members should work physically near to each other;
- each team member should have an agreed job title and job description.

6.5 Contract Problems

If work is contracted out, normal contract practices should be adopted but the following discussion serves to highlight certain important points. A detailed description of the work to be done and a cost estimate should be developed before contacting potential companies. It is useful to break down the project into a number of tasks for which manpower, equipment, material and overhead costs can be calculated. Allowances should be made for human underestimation of the duration of a task. A contingency factor for each task should also be added. It is sometimes useful to compare cost estimates with the money spent on previous similar projects. For preparation of 'calls for tender', 'requests for proposals' or contracts, the technical specification should be as detailed as possible, unambiguous and should be written by an expert and agreed on by interested parties within the organization before sending them out. 'Requests for proposals' are often used when the task cannot be well defined. The end product is often a report and hence they are best suited for the initial stages of a project. The exact details which form the content of a subsequent contract are usually negotiated, following an initial selection procedure. Calls for tender are usually more specific and hence the description of the requirements must be very specific, too. An organization which bids, does so knowing that a price is attached to the bid which if accepted, is usually adhered to unless later negotiations change the description of requirements.

Contractors can be selected on the basis of:

- the least expensive tender,
- the organization most likely to provide the desired product as specified in the contract,
- the reputation of the tendering organization,
- previous experience in the area,
- stability of the company,
- service facilities in the geographic area.

In order to establish the capabilities of the company the following can be examined:

- the management organization,
- the organizational procedures used throughout the company,
- standards used by the company to carry out work,
- controls and checks of work normally adopted,
- use of modern techniques such as structured programming, HIPO charts etc.,
- qualifications and experience of staff (specify the staff who will work on the project if feasible),
- examination of previous work.

A site visit to the company can be made, especially an unexpected one, to inspect the organization and the quality of the staff. For the evaluation of offers, the following should also be considered in addition to the companies capabilities:

- does the offer satisfy all the requirements?
- have the company deliberately underbid?
- summarize each offer to facilitate comparison,
- are the time constraints feasible?

Bids should be evaluated by a group of people (review panel) who understand the field of application. Controls should be built into the specification of a contract e.g. progress meetings, inspections, interim reports etc. Other factors to be specified include:

- subcontracting rules,
- testing criteria for product (benchmarks etc.),
- inspections or audits to be made at the suppliers works (if necessary).

Once a contractor has been selected, a single person should be nominated to communicate with the contractor. (It is usual to insist that the contractor do the same.) This person will control the execution of the contract through:

- regular progress meetings,
- keeping detailed documentation on the progress of the project,
- using an agreed 'change notification procedure',
- ensuring interim payments are only made following satisfactory progress,
- carrying out inspections etc.,
- ensuring that the product as delivered is satisfactory before authorizing payment.

6.6 Operational Problems

The introduction of a computer system requires an extremely methodological approach to the day-to-day running of the library or information service. Although the computer offers advantages in terms of power, speed etc. over manually maintained files, it also demands a very systematic approach. In order to achieve this, the use of a manual of procedures is recommended. This manual should contain a checklist of the way a particular task is to be carried out.

A related problem is that of systems maintenance. This involves the corrections of errors in the system, implementation of new versions and other changes. The amount of effort needed for this aspect of the system is often underestimated.

If a computer is used outside the library or information services, a procedure manual for the computer operators is essential to ensure that the regular tasks are carried out on schedule. Manuals should cover every aspect of the service, such as security procedures for dealing with classified information on magnetic tape and instructions for running computer programs etc.

A data-base administrator should be appointed. His function will be to arrange the safety and availability of data at all times and to ensure that data files are not corrupted (damaged).

The assignment of individual job titles and descriptions is also valuable so that each person has a clear understanding of his responsibilities.

7. CHEMICAL STRUCTURE INFORMATION SYSTEMS – A BRIEF INTRODUCTION

This chapter contains a brief introduction to input and storage of chemical structure information. This subject is one that has been comprehensively dealt with in the literature and several texts[34,35,36] are available.

The major ways of representing a chemical structure are the nomenclature, the molecular formula and the structure diagram. The problem of inputting the first two of these is no different from that of other information in character form. Retrieval of the desired chemical substance can be achieved using molecular formulae and nomenclature indexes, which are produced by computer using this information. The indexes to *Chemical Abstracts* show the variety of indexes which can be produced.

In order to store a structure diagram in the computer it is usually necessary to transfer the structure into symbols which can be directly input to the machine. The first problem is to transfer the two or three dimentional chemical structure diagram into a set of characters which can be manipulated by the computer. There are two basic approaches to storing information on chemical structures by means of a string of symbols.

The first is a topological representation, often known as a connection table or connectivity matrix which describes the atoms and their interconnections (bonds). Each type of atom and bond of a structure is given a symbol and individual atoms are given a number. A table of atom types and bond connections can be built up, which can then be stored in the computer. This system was adopted by Chemical Abstracts Service (CAS) for its chemical registry system.

The second approach is based on the use of linear notation. The best known of these is that devised by W.J.Wiswesser in the early 1950's. It comprises a string of alphanumeric characters which are used to describe a structure. Symbols are assigned to individual atoms or groups of atoms and bonds. For example, acylic atoms are given a single letter e.g. Fluorine 'F', Chlorine 'G'. A benzene ring is represented by 'R'. A fuller description can be found in Ash and Hyde[35].

Input of chemical structure information can be undertaken by entering the string of symbols corresponding to the structure. Linear notations are usually shorter than connection tables and hence can be encoded and input more quickly. On the other hand the connection table can be constructed by someone who is not a chemist, whereas the notation demands a knowledge of chemistry.

Alternative methods of input do exist. Zamora and Dayton[37] describe the use of a chemical typewriter for input of structure data at Chemical Abstracts Service. The system comprises a Varian 620i minicomputer together with twelve chemical typewriters which are specially modified IBM 735 Selectric typewriters with reverse index and a special typing element. Special characters such as ⫽ ‖ / \ are included. CAS also have interactive graphics devices for input. Another graphic method is to use optical scanning of structures drawn on a special grid.

Chemical structure information can be subjected to a number of error checks before being stored. Examples of checks are:

- use of correct symbols,
- comparison with molecular formula (input separately),
- validity of syntax.

An important feature of a chemical structure information system is the process of registration in which each unique structure is given a unique number. The difficult problem is to avoid registering duplicates. CAS uses a computer program to create a unique connection table from the one input which is then compared with all those already registered.

Chemical structure search packages are available. One example is the CROSSBOW system of Imperial Chemical Industries UK, which uses both Wiswesser Line notations and connection tables for different functions. Compounds having combinations of desired fragments or substructures can be retrieved from the computer and the respective structure diagrams can be displayed.

REFERENCES

1. Gibbs, D.C.C. *Glossary of Documentation Terms, Computer-User Terms.* AGARDograph No.182 – Part II.

2. Ralston, A.; Meek, C.L. (eds) *Encyclopaedia of Computer Science.* (First Edition) Petrocelli Charter, New York, 1976.

3. Kent, A.; Lancour, H.; Daily, J.E. (eds) *The Encyclopaedia of Library and Information Science.* Dekker, New York, (Vol.1–25).

4. – *The Application of Inexpensive Minicomputers to Information Work.* AGARD Lecture Series No.92, 1978.

5. Mao, A.R. *Basic Digital Computer Concepts, Principles of Avionics Computer Systems.* AGARDograph No.183, pp.3-29.

6. – *Auerbach Dataworld.* Auerbach publishers Inc., 6560 North Park Drive, Pennsaiken, N.J. 08109, USA, 1963. A four volume loose-leaf series.

7. Coffman, E.G.; Denning, P.J. *Operating Systems Theory.* (First Edition) Prentice Hall, 1973.

8. Groetz, M.A. *The Software Products Industry: Its Future and Promise*, Computerworld, pp.9-19, October 16, 1978.

9. – *GRIPS, A General Relation-based Information Processing System – Documentation (Einfuhrung).* DIMDI, Cologne, Federal Republic of Germany.

10. Daneluik, F.A. *The IDRC Minicomputer-based Bibliographic Information System.* AGARD Lecture Series No.92, pp.4.1-4.9, 1978.

11. Grosch, A.N. *Commercial Data Base Management (DBMS) Software.* AGARD Lecture Series No.92, pp.8.1-8.10, 1978.

12. Martin, J. *Computer Data-base Organization.* Prentice-Hall, 1975.

13. Date, C.J. *An Introduction to Database Systems.* Addison-Wesley, 1975.

14. Martin, J. *Principles of Data-base Management.* Prentice-Hall, 1976.

15. Lynch, M.F. *Computer-based Information Services in Science and Technology – Principles and Techniques.* (First Edition) Peter Perigrinus, 1974.

16. – *7-bit Coded Character Set for Information Processing Interchange.* ISO – 646, International Organization for Standardization, Geneva, Switzerland, 1973.

17. Dunne, J.C.; Howe, R.H.; Pocock, O.E.; Edwards, T. *DRIC Manual for Abstracting and Subject Indexing.* DRIC-BR-44600, Defence Research Information Centre, UK, January 1975.

18. UNISIST *Reference Manual for Machine-readable Bibliographic Descriptions.* (First Edition) UNESCO, Paris, 1975.

19. – ISBD(G) General, First Edition, International Federation of Library Association, International Office for UBC, London, 1977. ISBN 0-903043-18-1.

20. Hines, T.C. *Minimizing Input Effort for Computer-based Information Systems, a Case Study Approach.* Spec. Libr., Vol.67 (8), pp.377-381, 1976.

21. Hart, G.W. — *The Application of Inexpensive Computers to Information Work.* AGARD Lecture Series No.92, pp.5.1-5.17, 1975.

22. – — *R and T Work Unit System Data Input Manual.* United States Defense Documentation Center, Defense Supply Agency, December, 1968.

23. Axner, D.E., Keagan, F.H. — *Alphanumeric Display Terminal Survey*, Datamation, Vol.24 (6), pp.183-192, June 1978.

24. – — *Alphanumeric Character Sets for Optical Character Recognition.* ISO Recommendation R1073, International Organization for Standardization, Geneva, Switzerland.

25. – — *Printing Specifications for Optical Character Recognition.* ISO Recommendation R1831, International Organization for Standardization, Geneva, Switzerland, November 1971.

26. Haritatos, F.N., Petruzelli, J.D. — *Advancements in Retrieval Technology as Related to Information Systems*, AGARD Conference Proceedings No.207, pp.14.1-14.7, October 1976.

27. Karten, H.A. — *Solid-state 2305-Type Unit a 'First'*, Computerworld, Vol.XII, No.39, pp.1, 6, September 25, 1978.

28. Grosch, A.N. — *Selection of Minicomputer Systems for Bibliographic Applications*, AGARD Lecture Series No.92, pp.2.1-2.8, 1978.

29. McIvor, R.A. — *The Application of Inexpensive Minicomputers to Information Work*, AGARD Lecture Series No.92, pp.1.1-1.5, 1978.

30. Katzan, H., Jr — *Systems Design and Documentation.* (First Edition) Van Nostrand, New York, 1976.

31. Gildersleeve, T.R. — *Data Processing Project Management.* (First Edition) Van Nostrand, New York, 1974.

32. Cleland, D.I., King, W.R. — *Systems Analysis and Project Management.* McGraw Hill, New York, 1975.

33. Bingham, J.E., Davies, G.W.P. — *A Handbook of Systems Analysis.* (Second Edition) Macmillan, London, 1978.

34. Lynch, M.F., Harrison, J.M., Town, W.G., Ash, J.E. — *Computer Handling of Chemical Structure Information.* McDonald, London and Elsevier, American, 1971.

35. Ash, J.E., Hyde, E. (eds) — *Chemical Information Systems.* (First Edition) Ellis Horwood Ltd, Chichester, 1975.

36. Davis, C.E., Rush, J.E. — *Information Retrieval and Documenation in Chemistry.* (First Edition) Greenwood Press, Westport Connecticut, 1974.

37. Zamora, A., Dayton, D.L. — *The Chemical Abstracts Service Chemical Resigtry System V Structure Input and Editing.* J. Chem. Info. Comp. Sci., Vol.16 (4), pp.219-222, 1976.

ACKNOWLEDGEMENTS

The author would like to express his appreciation and thanks to all those who sent useful material for this publication, especially Mr Chandler of NASA, Mr Kurzwelly of DIMDI, Federal Republic of Germany, Dr McIvor of the Canadian National Defence Organization, Mr Ryerson of the United States National Technical Information Service, Mr Salmon of ONERA, France, Mr Sauter of the United States Defense Documentation Center, Ms Edwards of the International Development Research Centre in Canada, Mr Clague of INSPEC, UK, and Mr Hart of the United Kingdom Defence Research Information Centre. I would also like to thank IBM Belgium, Sait Electronics Brussels (agents for Datapoint equipment) and Nixdorf Computer AG for supplying useful material on equipment, and Mr Garth Davies for his valuable advice.

APPENDIX A

AN EXAMPLE OF ON-LINE DATA ENTRY AT THE LIBRARY OF THE INTERNATIONAL DEVELOPMENT RESEARCH CENTRE (CANADA)

The following dialogue shows how a record of an item is to be ordered, catalogued, and entered into a data base, via an on-line terminal. When the item subsequently arrives in the library, the cataloguers further process the information to produce an expanded bibliographic description. Finally an abstract is prepared.

```
VALID FUNCTIONS ARE:
      ENTRY      }
      INDEX      }       options for interacting with
      LISTFORMAT }       the data base.
      MODIFY     }       'entry' is selected for a new
      PRINT      }       item to be input.
      QUERY      }
      EXIT       }

WHICH FUNCTION DO YOU WANT? entry ←— item to be ordered.
PROGRAM LOADED WITH LIB = P
ENTRY A.01.04 LOGON MON, JUL 10, 1978,  3:19 PM
PLEASE ENTER DATA BASE NAME OR EXIT
? process
PLEASE SELECT BIBLIOGRAPHIC LEVEL ? a
ISN = 29008
PERSONAL AUTHOR-ANALYTIC
? Daniels, W.C.
PERSONAL AUTHOR-ANALYTIC
? MacCormac, C.W.
PERSONAL AUTHOR-ANALYTIC
? Hulse, J.H.
PERSONAL AUTHOR-ANALYTIC
?
CORPORATE AUTHOR-ANALYTIC
CORPORATE AUTHOR NAME - ANALYTIC
? 000711
CORPORATE AUTHOR CODE-ANALYTIC
? 000711
CORPORATE AUTHOR PART-ANALYTIC
? Agriculture, Food and Nutrition Sciences Division
MORE SUBFIELDS(Y/N) ? n
AFFILIATION-ANALYTIC
AFFILIATION CODE-ANALYTIC
?
AFFILIATION PART-ANALYTIC
?
TITLE-ANALYTIC
? Collaboration in agricultural research
TITLE-ANALYTIC
?                            ←—
PAGE NOS.-ANALYTIC                at this point the system checks
? p. 9-18                         to see if the title is already
PERSONAL AUTHOR-MONOGRAPH         in the data base (check for
?                                 duplicates)

CORPORATE AUTHOR-MONOGRAPH
CORPORATE AUTHOR NAME-MONOGRAPH
?
CORPORATE AUTHOR CODE-MONOGRAPH
```

```
?
CORPORATE AUTHOR PART-MONOGRAPH
?
MORE SUBFIELDS(Y/N) ?
AFFILIATION-MONOGRAPH
AFFILIATION CODE-MONOGRAPH
?
AFFILIATION PART-MONOGRAPH
?
TITLE-MONOGRAPH
?
COLLATION-MONOGRAPH
?
VENDOR CODE
?
NON-VENDOR
?
STATUS CODE
?  pprint
STATUS CODE DATE (DD/MM/YY)
? ++
ESTIMATED PRICE
?
ORDER DATE (DD/MM/YY)
?
MATERIAL DUE DATE (DD/MM/YY)
?
DATE RECEIVED (DD/MM/YY)
?
CITATION SOURCE
?
REQUESTED BY:
REQUESTOR NAME
?
REQUESTOR'S DIVISION
?
MORE SUBFIELDS(Y/N) ?
NUMBER OF COPIES, FORMAT
?
ALLOCATION AND COMMITMENT NUMBER
?
ACCOUNT NUMBER
?
ACQUISITIONS NOTES
?
DIRE: ADDRESS (5 LINES)
DIRE--NAME LINE 1
?
DIRE--NAME LINE 2
?
DIRE--ADDRESS LINE 1
?
DIRE--ADDRESS LINE 2
?
DIRE--ADDRESS LINE 3
?
LANGUAGES OF TEXT
? ENGL
CONFERENCE
CONFERENCE NAME
?
```

```
CONFERENCE PLACE
?
CONFERENCE DATE
?
CONFERENCE COUNTRY
?
MORE SUBFIELDS(Y/N) ?
EDITION
?
PUBLISHER
PUBLISHER NAME
?
PLACE
?
COUNTRY
?
ISBN
?
DATE OF IMPRINT/ISSUE
? Aug. 1977
LIBRARY OF CONGRESS CARD NUMBER
?
CANADIANA CONTROL NUMBER
?
```

At this stage the item has arrived in the library and its 'status' is changed to 'inpro' i.e. in process.

```
VALID FUNCTIONS ARE:
      ENTRY
      INDEX
      LISTFORMAT
      MODIFY
      PRINT
      QUERY
      EXIT
```

modify function chosen so that data can be modified

```
WHICH FUNCTION DO YOU WANT? modify
MODIFY VERSION A.02.00 LOGON MON, JUL 10, 1978,  3:31 PM
PLEASE SELECT DATA BASE OR EXIT
?biblio
PLEASE ENTER FUNCTION TO SELECT RECORD
?isn=29008
ISN = 29008
PLEASE SELECT FUNCTION
?l all
29008            STATUS: inpro
CHANGD:   CHANGI:
LANTEX: ENGL    DATEC:       DATEI: Aug. 1977
PAUTHA: Daniels, W.C.    MacCormac, C.W.    Hulse, J.H.
CORPA : 000711000711Agriculture, Food and Nutrition Sciences Division
TITLEA: Collaboration in agricultural research
PAGENA: p. 9-18
FACET :
ABST  :

PLEASE SELECT FUNCTION
?tra a174 to a144
CORPORATE AUTHOR-ANALYTIC
000711000711Agriculture, Food and Nutrition Sciences Division
CORPORATE AUTHOR CODE-ANALYTIC
000711
```

```
** FIELD TRANSFERRED
PLEASE SELECT FUNCTION
?tra a175 to a145
CORPORATE AUTHOR-ANALYTIC
000711Agriculture, Food and Nutrition Sciences Division
CORPORATE AUTHOR PART-ANALYTIC
Agriculture, Food and Nutrition Sciences Division
AFFILIATION-ANALYTIC
000711
** FIELD TRANSFERRED
PLEASE SELECT FUNCTION
PLEASE SELECT FUNCTION
?del a171
CORPORATE AUTHOR-ANALYTIC
000711
CORPORATE AUTHOR NAME - ANALYTIC
000711
** FIELD CORPNA DELETED
PLEASE SELECT FUNCTION
?l a170
** FIELD CORPA  IS NOT DEFINED
PLEASE SELECT FUNCTION
?l a174
** FIELD CORPCA IS NOT DEFINED
PLEASE SELECT FUNCTION
?l a140
AFFILIATION-ANALYTIC
000711Agriculture, Food and Nutrition Sciences Division
PLEASE SELECT FUNCTION
?add datec
DATE OF CONTENT
?1977
PLEASE SELECT FUNCTION
?cha status
STATUS CODE
inpro
PLEASE ENTER CHANGE
?inpro%proof
PLEASE ENTER CHANGE
?
proof
PLEASE SELECT FUNCTION
?add change
LAST CHANGE
LAST CHANGE DATE
?++
LAST CHANGE INITIALS
?cs
PLEASE SELECT FUNCTION
```

Following this the item is abstracted etc.

APPENDIX B

SOME EXAMPLES OF INPUT FORMATS, FORMS, HANDLING OF SPECIAL SYMBOLS AND ENTRIES IN A DATA DICTIONARY

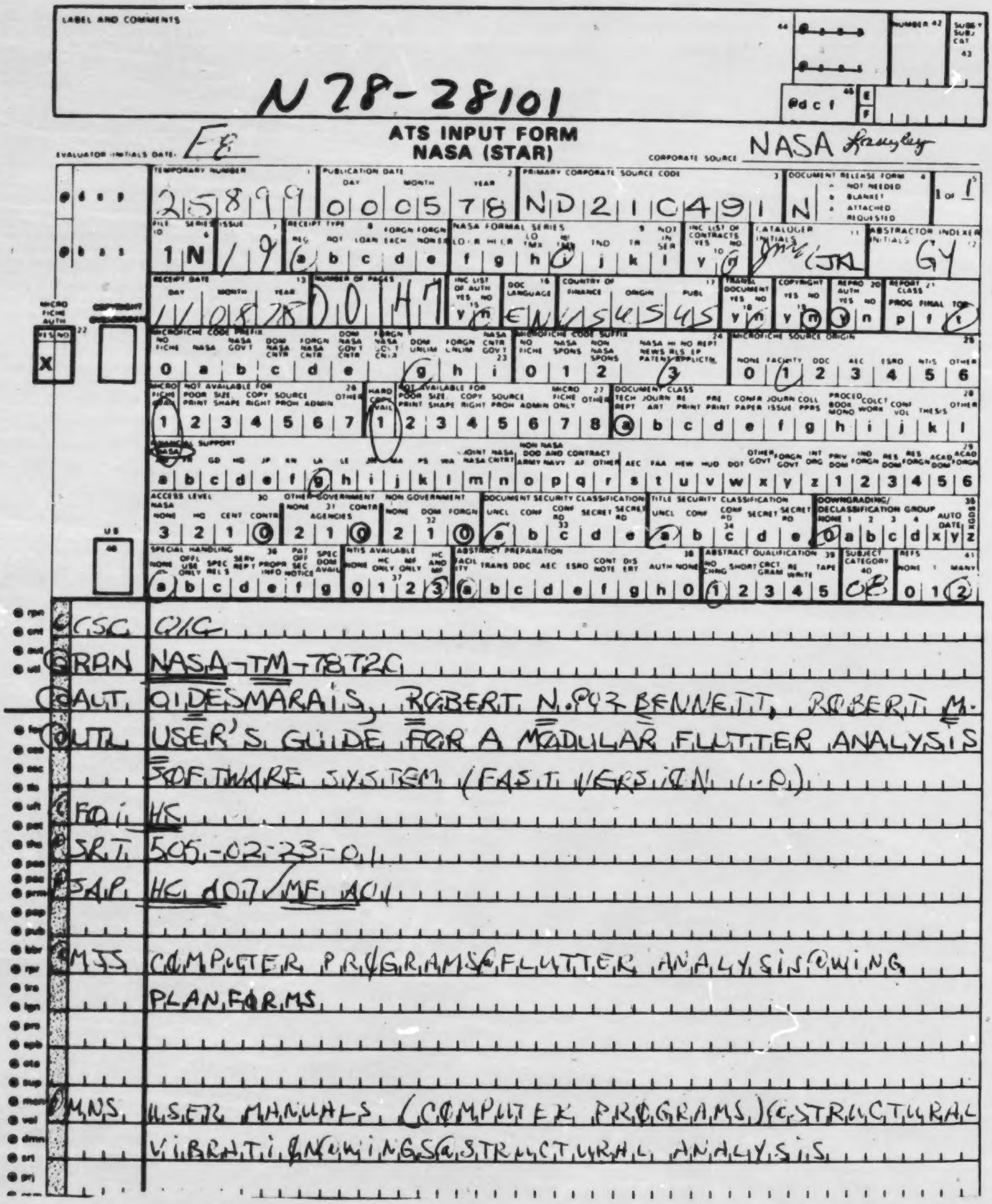
Figure B1 shows an input form used for the NASA STAR data base. Much of the information is indicated by marking the appropriate position on the page. Some information e.g. publication data, is in fixed field format.

Another example of a fixed field format is that used for the Research and Technology Work Unit Information System (WUIS) at the US Defense Documentation Center (DDC).

Figure B2 shows the data encoded on a single punched card.

Figure B3 shows some of the rules for handling special symbols.

Figure B4 shows an example taken from the data dictionary (format description), showing the instructions for encoding a title. The label, (tag or field number) is "11", the name is 'Title' and it has a length of 160 characters which must be alphanumeric.



LABEL AND COMMENTS

N78-28101

Pdcf

ATS INPUT FORM
NASA (STAR)

EVALUATOR INITIALS DATE: EE

CORPORATE SOURCE: NASA Langley

TEMPORARY NUMBER (1): 25899

PUBLICATION DATE (2): DAY 00 MONTH 05 YEAR 78

PRIMARY CORPORATE SOURCE CODE (3): ND21C491

DOCUMENT RELEASE FORM (4): N (a NOT NEEDED, b BLANKET, c ATTACHED, REQUESTED)

(5): 1 of 1

FILE ID / SERIES (6): 1N

ISSUE (7): 19

RECEIPT TYPE (8): REG [✓ a] NOT b LOAN c FORGN EACH d FORGN NONEX e

NASA FORMAL SERIES (9): f g h [✓ i] j k l

INC LIST OF CONTRACTS (10): YES y NO [✓ n]

CATALOGER INITIALS (11): JKL

ABSTRACTOR INDEXER INITIALS (12): GY

RECEIPT DATE (13): DAY 11 MONTH 08 YEAR 78

NUMBER OF PAGES (14): 0047

INC LIST OF AUTH (15): YES y NO [✓ n]

DOC LANGUAGE (16): EN

COUNTRY OF (17): FINANCE US ORIGIN 45 PUBL 45

TRANSL DOCUMENT (18): YES y NO [✓ n]

COPYRIGHT (19): YES y NO [✓ n]

REPRO AUTH (20): YES [✓ y] NO n

REPORT CLASS (21): PROG p FINAL f TOP [✓ t]

MICRO FICHE AUTH (22): YES ☒ NO ☐

COPYRIGHT

MICROFICHE CODE PREFIX (23): NO FICHE 0, NASA a, NASA GOVT b, DOM NASA CNTR c, FORGN NASA CNTR d, DOM NASA GOVT CNTR e, FORGN NASA GOVT CNTR [✓ g], DOM UNLIM h, FORGN UNLIM / NASA CNTR GOVT i

MICROFICHE CODE SUFFIX (24): NO FICHE 0, NASA SPONS 1, NON NASA SPONS 2, NASA HI NO REPT NEWS RLS EP PATENT APPLICTN [✓ 3]

MICROFICHE SOURCE ORIGIN (25): NONE 0, FACILITY [✓ 1], DDC 2, AEC 3, ESRO 4, NTIS 5, OTHER 6

MICRO FICHE AVAIL (26): [✓ 1]; NOT AVAILABLE FOR: POOR PRINT 2, SIZE SHAPE 3, COPY RIGHT 4, SOURCE PROM 5, ADMIN 6, OTHER 7

HARD COPY AVAIL (27): [✓ 1]; NOT AVAILABLE FOR: POOR PRINT 2, SIZE SHAPE 3, COPY RIGHT 4, SOURCE PROM 5, ADMIN 6, MICRO FICHE ONLY 7, OTHER 8

DOCUMENT CLASS (28): TECH REPT [✓ a], JOURN ART b, RE PRINT c, PRE PRINT d, CONFR PAPER e, JOURN ISSUE f, COLL PPRS g, PROCED BOOK MONO h, COLCT WORK i, CONF VOL j, THESIS k, OTHER l

FINANCIAL SUPPORT (29): [✓ NASA] a b c d e f [✓ g] h i j k l m n o p q r s t u v w x y z 1 2 3 4 5 6

ACCESS LEVEL (30): NASA NONE 3, HQ 2, CENT 1, CONTR [✓ 0]

OTHER GOVERNMENT (31): NONE 2, AGENCIES 1, CONTR [✓ 0]

NON GOVERNMENT (32): NONE 2, DOM 1, FORGN [✓ 0]

DOCUMENT SECURITY CLASSIFICATION (33): UNCL [✓ a], CONF b, CONF RD c, SECRET d, SECRET RD e

TITLE SECURITY CLASSIFICATION (34): UNCL [✓ a], CONF b, CONF RD c, SECRET d, SECRET RD e

DOWNGRADING/DECLASSIFICATION GROUP (35): NONE [✓ 0], 1 a, 2 b, 3 c, 4 d, AUTO DATE x y z

SPECIAL HANDLING (36): NONE [✓ a], OFFL USE ONLY b, SPEC RELS c, SERV REPT d, PROPR INFO e, PAT OFF SEC NOTICE f, SPEC DOM AVAIL g

NTIS AVAILABLE (37): NONE 0, HC ONLY 1, MF ONLY 2, HC AND MF [✓ 3]

ABSTRACT PREPARATION (38): FACILITY [✓ a], TRANS b, DDC c, AEC d, ESRO e, CONT NOTE f, DIS ERT g, AUTH h, NONE 0

ABSTRACT QUALIFICATION (39): NO CHNG [✓ 1], SHORT 2, CRCT GRAM 3, RE WRITE 4, TAPE 5

SUBJECT CATEGORY (40): 08

REFS (41): NONE 0, 1 1, MANY [✓ 2]

US 40

CSC: OIC

RPN: NASA-TM-78720

ALT: OIDESMARAIS, ROBERT N.@2 BENNETT, ROBERT M.

UTL: USER'S GUIDE FOR A MODULAR FLUTTER ANALYSIS SOFTWARE SYSTEM (FAST VERSION 1.0)

FQI: HC

SRT: 505-02-23-01

SAP: HC A07/MF A01

MJS: COMPUTER PROGRAMS@FLUTTER ANALYSIS@WING PLANFORMS

MNS: USER MANUALS (COMPUTER PROGRAMS)@STRUCTURAL VIBRATION@WINGS@STRUCTURAL ANALYSIS

Fig.B1 Data collection form for the NASA (STAR) data base

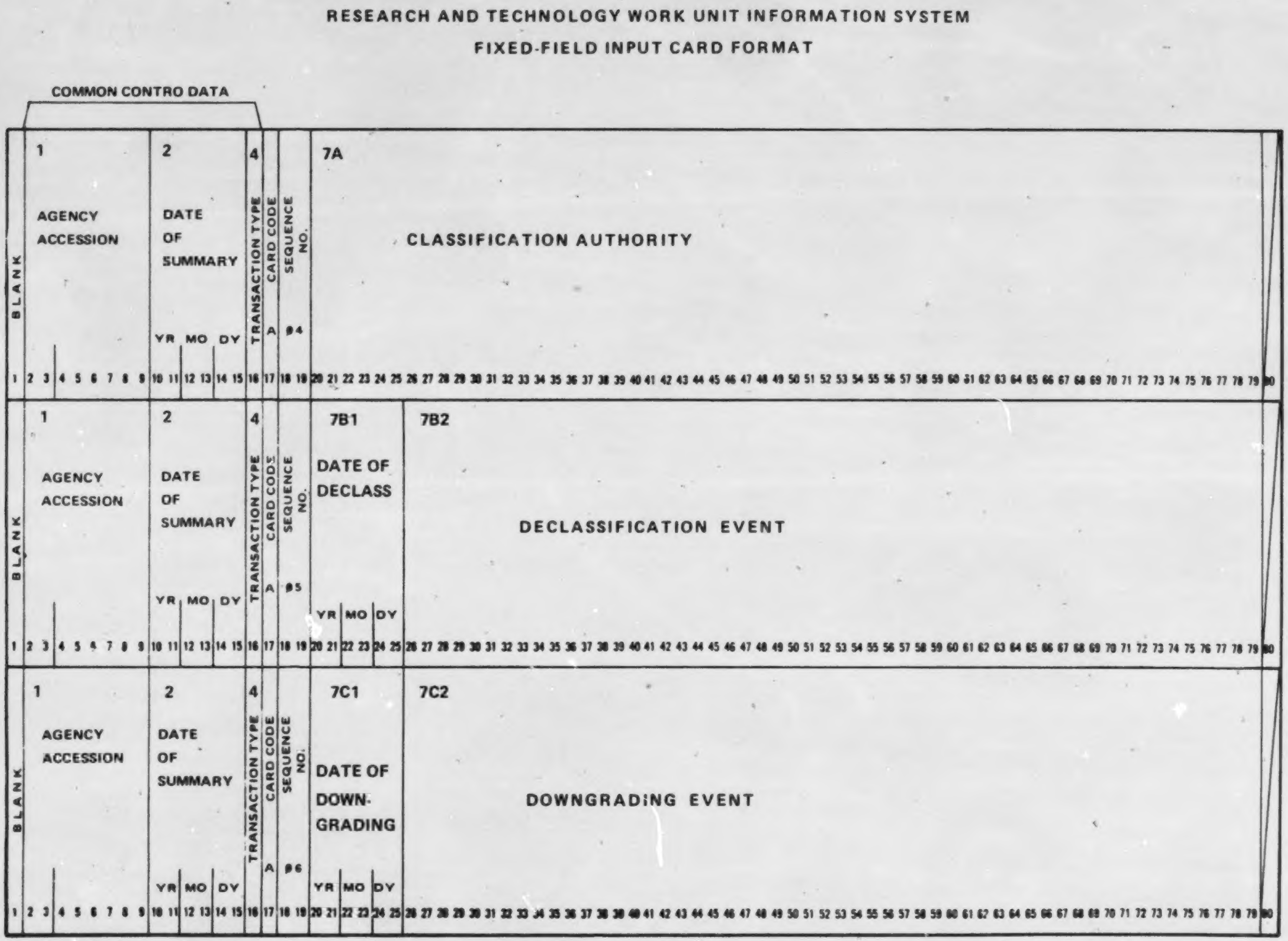

RESEARCH AND TECHNOLOGY WORK UNIT INFORMATION SYSTEM
FIXED-FIELD INPUT CARD FORMAT
COMMON CONTRO DATA
BLANK
1
AGENCY ACCESSION
2
DATE OF SUMMARY
YR
MO
DY
4
TRANSACTION TYPE
CARD CODE
A
SEQUENCE NO.
04
7A
CLASSIFICATION AUTHORITY
05
7B1
DATE OF DECLASS
7B2
DECLASSIFICATION EVENT
06
7C1
DATE OF DOWN-GRADING
7C2
DOWNGRADING EVENT
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

Fig.B2 Part of the fixed field input format used at the US Defense Documentation Center

VERBALIZING FOR MACHINABILITY*

SPECIAL SYMBOLS

$\left.\begin{matrix}\cong\\ \approx\end{matrix}\right\}$ *use* approximately

$\rightarrow\left\{\begin{matrix}\text{use yields (chemistry)}\\ \text{use approaches limit of (mathematics)}\end{matrix}\right.$

# *use* no.

& *use* and *in titles, abstracts, and annotations*

∞ *use* infinity

λ *use* wavelength *(electronics and physics)*
use lambda *(all other)*

Ω *use* ohms *(electricity and electronics)*
use omega (all other)

ϕ *use* phase *(electricity and electronics)*
use phi *(all other)*

Similarly, spell out or show by acceptable alphanumeric characters increment, varies as, therefore, differential of, variation of, integral, sum, benzene ring, thunderstorm, male, female, fixed star, *etc.*

SQUARE

cm^2 *use* sq cm
ft^2 *use* sq ft
m^2 *use* sq m

SQUARE ROOT

$\left.\begin{matrix}\sqrt{a-b}\\ (a-b)^{1/2}\end{matrix}\right\}$ *use* square root of (a-b)

SUBSCRIPTS

V_1 *use V sub 1*

B_5 *use* B *(omit the 5, which is the atomic number of boron)*
See also CHEMICALS

SUPERSCRIPTS

H^+ *use* H(+)

SO_4^{--} *use* SO4(–)

V^{5+} *use* V(5+)

U^{234} *use* U234

B^{10} *use* B10

$O^{18}(p,n)N^{15}$ *use* O18(p,n)N15

d_{23}^{25} *use* density at 23 deg F referred to water at 25 deg F

n_D^{20} *use* index of refraction for 20 deg F and sodium light
See also CUBIC, EXPONENTS, SQUARE

UNDERSCORING

Do not use underscoring
Escherichia coli *use* Escherichia coli
to set off special terms use single quotes; e.g.,
the term quasar *use* the term 'quasar'

Fig.B3 Some of the rules for handling special symbols at the United States Defense Documentation Center

Field 11 **Title (160 Alpha/Numeric)**

Instructions

A brief descriptive title for the work unit is mandatory at all times in a WUIS record. An entry is mandatory on all NEW transactions. The field may not be deleted by a modification transaction. The first character of the title field in the card or tape record is a single-character code for the security classification of the title (either U, C, S, or T) When preparing cards or card images do not leave a space between the code and the first character of the title itself. (Items classified Top Secret must be forwarded to NSA, and will not be accepted by DDC.)

Edit/Audit

a. The field may never be blank or deleted.

b. The first character of the title must be U, C, or S for entry in the WUIS by DDC.

c. The security classification of the title must be equal to or less than the Summary Security, Field 5.

Disposition

a. A transaction is not releasable if Field 11 is blank on a NEW transaction or deleted by a modification.

b. A transaction is not releasable if the title classification is other than U, C, or S or if the title classification is higher than Field 5.

Fig.B4 A description of the data element 'title' in the US Defense Documentation Center data base (an item from a data dictionary)

APPENDIX C

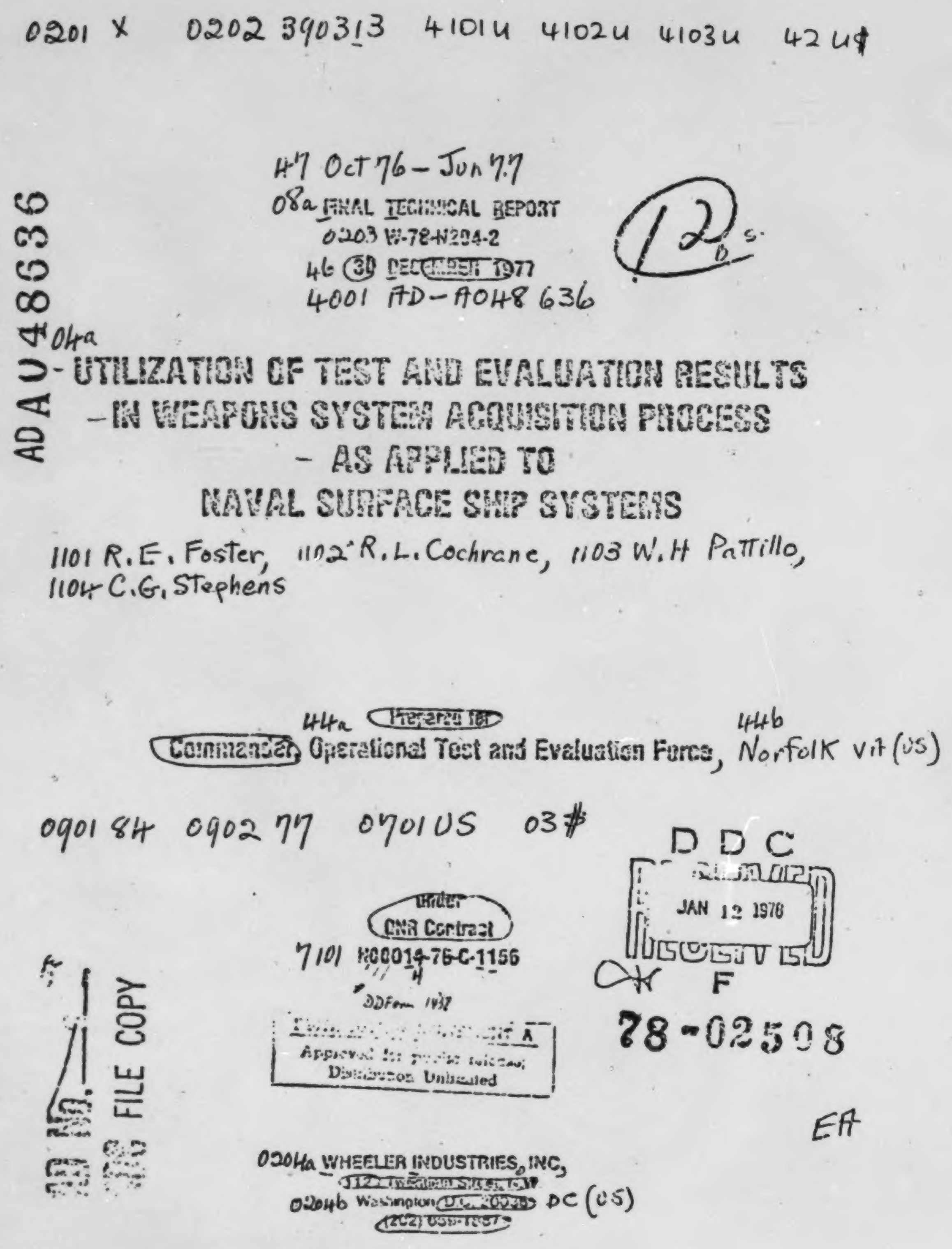
PROCESSING OF BIBLIOGRAPHIC INFORMATION AT THE CANADIAN NATIONAL DEFENCE INFORMATION CENTRE

Figure C1 shows the title page from a report to be included in the system.

Figure C2 shows the print-out from the computer used for proof-reading purposes.

Figure C3 shows a temporary catalogue card and a second card used for loan control.

The examples shown in Figure C2 and Figure C3 are produced by the computer, which also produces the library catalogue on microfilm.



0201 X 0202 390313 4101U 4102U 4103U 42 U#

47 Oct 76 – Jun 77
08a FINAL TECHNICAL REPORT
0203 W-78-N204-2
46 (30) DECEMBER 1977
4001 AD-A048 636

12

AD A048636

04a – UTILIZATION OF TEST AND EVALUATION RESULTS
– IN WEAPONS SYSTEM ACQUISITION PROCESS
– AS APPLIED TO
NAVAL SURFACE SHIP SYSTEMS

1101 R.E. Foster, 1102 R.L. Cochrane, 1103 W.H Pattillo,
1104 C.G. Stephens

44a Prepared for
Commander, Operational Test and Evaluation Force, 44b Norfolk VA (US)

0901 84 0902 77 0701US 03#

D D C
JAN 12 1978
F

under
ONR Contract
7101 N00014-76-C-1156

DDForm 1437

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

78-02508

EA

FILE COPY

020Ha WHEELER INDUSTRIES, INC,
1120 [illegible]
0204b Washington D.C. 20036 DC (US)
(202) [illegible]

Fig.C1 The title page from a report which is to be catalogued

```
       01 78-02508 41 U U U 42 U$    90 HF 4      02                 390313
 W-78-N204-2              Wheeler Industries Inc, Washington DC (US)

          04 UTILIZATION OF TEST AND EVALUATION RESULTS IN WEAPONS SYSTEM ACQUI
SITION PROCESS AS APPLIED TO NAVAL SURFACE SHIP SYSTEMS
                                                           21

                                                                    46 Dec 77
            35                                                 71 N00014-76-C-1156
                                         40 AD-A048 636





BIB002 11 Foster, R.E.                              * Cochrane, R.L.
          Pattillo, W.H.                            * Stephens, C.G.
88  1 KBS            * NTIS
   090878
47 Oct 76 - Jun 77
08 Final Technical Report

06

30

31

36                  *
38
72
44 Operational Test and Evaluation Force, Norfolk VA (US)

09 84            * 77                *
07 US               *
22               *         *

                                              *
BIB003
23               *         *

03 $
20
97
29

05

99

51 Gun fire control                          * Procurement
   Sonar equipment                           * Marine propulsion
   Gas turbine rotors                        * Acquisition
   Systems engineering                       * Reliability
   Destroyers                                *
                                             *
                                             *
                                             *
                                             *
```

Fig.C2 Print out of the bibliographic details of a report (see Figure C1) for proof-reading purposes (continued)

```
    *
 BIB004
 52                                        *
                                           *

 81 1505    * 1310    *        *        *        *
 12                  *                *                *
                     *                *                *
                     *                *                *
                     *                *                *
                     *                *                *
 50 The report documents the results of a comprehensive review of four surface s
hip subsystem developments, acquisition, test and evaluation, with emphasis on t
he use of test and evaluation results in the acquisition process decisions. The
systems studied: (1) DD 963 Class Propulsion System; (2) CGN-36 Class and CGN-38
 Class Combat System Integration Effort; (3) MK-86 Gun Fire Control System; and
(4) SQS-56 Sonar.
```

Figure C2 (concluded)

```
      01 78-02508 41 U U U 42 U$     90 HF 4
02                 390313     W-78-N204-2
     Wheeler Industries Inc, Washington DC (US)


04 UTILIZATION OF TEST AND EVALUATION RESULTS IN
WEAPONS SYSTEM ACQUISITION PROCESS AS APPLIED TO N
AVAL SURFACE SHIP SYSTEMS

21


                         46 Dec 77
     35
71 N00014-76-C-1156
     40 AD-A048 636
```

```
      01 78-02508 41 U U U 42 U$     90 HF 4
02                 390313     W-78-N204-2
     Wheeler Industries Inc, Washington DC (US)


04 UTILIZATION OF TEST AND EVALUATION RESULTS IN
WEAPONS SYSTEM ACQUISITION PROCESS AS APPLIED TO N
AVAL SURFACE SHIP SYSTEMS

21


                         46 Dec 77
     35
71 N00014-76-C-1156
     40 AD-A048 636
```

Fig.C3 Catalogue cards produced by the computer

Section 5

MECHANIZATION SYSTEMS AND OPERATIONS

by
Victor J.Rogers
Technology Reports Centre
Department of Industry
Orpington, Kent, UK

ABSTRACT

This Section provides a working basis for setting up a computer system for indexing, processing and disseminating scientific and technical information mainly in the form of bibliographic references to research report literature. Following brief references to alternative forms of mechanisation, such as the tape typewriter, the main emphasis is upon the in-house computer and its uses for data capture, text handling, good-quality printing and information retrieval. Methods are described for setting up, storing and exploiting databases, and the procedures for creating an announcement journal are defined in detail.

The practical examples which are used to illustrate the work are based upon experience gained at the Technology Reports Centre (TRC) of the UK Department of Industry and the Defence Research Information Centre (DRIC) of the UK Ministry of Defence.

CONTENTS

Page

1. INTRODUCTION 57

2. BASIC MECHANISATION 57
   2.1 Tape Typewriters 57
   2.2 Computerised Indexes via the Tape Typewriter 57
   2.3 Computer-based Data Store 58

3. THE DATA BASE 58
   3.1 Storage of Data Within the Computer 59
   3.2 The Computerised Thesaurus 60

4. THE IN-HOUSE COMPUTER 62
   4.1 Initial Preparations 62
   4.2 Input File Structures 65
   4.3 Storage File Structures 69
   4.4 Output File Structure 69
   4.5 Data Preparation Systems – General 69
   4.6 Key-to-disc Input from Terminals Linked to an In-house Computer 70
   4.7 The Data Fields 73
   4.8 Separator Codes 75
   4.9 Keying-in, Validating and Correcting Data 75
   4.10 Operator Aids 76
   4.11 Indexes 76
   4.12 Key-to-disc Input from Free-standing Terminal Stations 80
   4.13 Key-to-tape Input Systems 81
   4.14 Input from External Sources 81
   4.15 Direct Printing via the Computer 81
   4.16 Printers 81
   4.17 The Remote Printing of Computer Output 87
   4.18 The Layout of Printed Pages 87
   4.19 Production Scheduling 91
   4.20 File Security 92
   4.21 The Handling of Security Classified Information 92
   4.22 Housekeeping and Stock Control 92
   4.23 Forms Design 94
   4.24 Computer Configurations 94

5. THE USE OF MINI-COMPUTERS IN THE PROCESSING OF RESEARCH REPORT LITERATURE 94
   5.1 Program Aspects 96
   5.2 Non-bibliographic Applications 96

6. COMPUTER BUREAU OPERATIONS 96
   6.1 Setting Up the Bureau Operation 96
   6.2 The Advantages of Bureau Operation 97
   6.3 The Disadvantages of Bureau Operation 97

ACKNOWLEDGEMENTS 97

BIBLIOGRAPHY 98

GLOSSARY OF TERMS 99

LIST OF FIGURES

		Page
Figure 1	Single-column page of text	58
Figure 2	Double-column page of text	59
Figure 3	Extract from subject index	60
Figure 4	Data base structure	61
Figure 5	Main bibliographic record	62
Figure 6	Page from the EJC Thesaurus	63
Figure 7	Thesaurus update	64
Figure 8	A typical organisational structure	66
Figure 9	A typical abstract	67
Figure 10	Bibliographic record displayed with control codes	67
Figure 11	Parallel file system	68
Figure 12	Typical storage file structure	70
Figure 13	Output file structure	70
Figure 14	A summary of the data preparation and abstract journal production systems at TRC	71
Figure 15	Summary of in-house computer system	72
Figure 16	Document processing sheet	74
Figure 17	A typical journal index system	77
Figure 18	Sample of a contents page	78
Figure 19	Creation and update of data base files	79
Figure 20	Harmonisation of data base files	82
Figure 21	The 96-character set used by the BCL M18 computer	83
Figure 22	A typical 96-character set	84
Figure 23	The Dataproducts 2230 drum printer	85
Figure 24	The Tally 3000-series matrix printer	85
Figure 25	The Diablo 1620 petal printer	86
Figure 26	The Versatec electrostatic printer	86
Figure 27	Computer typeset abstract journal	88
Figure 28	The Linotron 202 computer-controlled photo-typesetter	89
Figure 29	Sample indexes from R & D Abstracts	91
Figure 30	A typical reports file record	93
Figure 31	A typical customer file record	93
Figure 32	A typical transactions file record	93
Figure 33	The GEC 4080 computer at DRIC	95

1. INTRODUCTION

Research and development in the realms of science and technology give rise to a flood of reports and special literature which is generally created at considerable cost. Modern techniques for dealing with this documentation in ways which will benefit the widest possible range of users rely increasingly upon mechanical and electronic aids, and it is evident that the computer is the most significant element in systems designed for processing scientific and technical information (STI). This applies particularly to the means used for providing existing and potential users of research report literature with lists of the holdings of information centres and special libraries, and it is true to say that efficient and comprehensive exploitation of modern data bases would be quite impracticable without the computer.

Computer manufacturers and specialised software houses offer a variety of systems which can be employed in mechanised information handling processes so that, when planning the computerisation of an existing manual system, it is wise to consult hardware manufacturers and organisations with experience of writing relevant software. If it is found practicable to purchase a commercial software package, i.e. a ready-made set of working programs, it is generally necessary to make significant additions to the system in order that the requirements of the local situation will be fully met. This section of the Manual describes the steps to be taken in order to build up a computer-based system designed both to handle the collection and processing of bibliographic references to scientific and technical report literature, and also to disseminate the information mainly in the form of announcement journals and similar accession lists. The recommended systems are capable of providing information retrieval by direct access from remote computer terminals.

The practical examples which are used to illustrate the work are based upon experience gained at the Technology Reports Centre (TRC) of the UK Department of Industry and the Defence Research Information Centre (DRIC) of the UK Ministry of Defence.

2. BASIC MECHANISATION

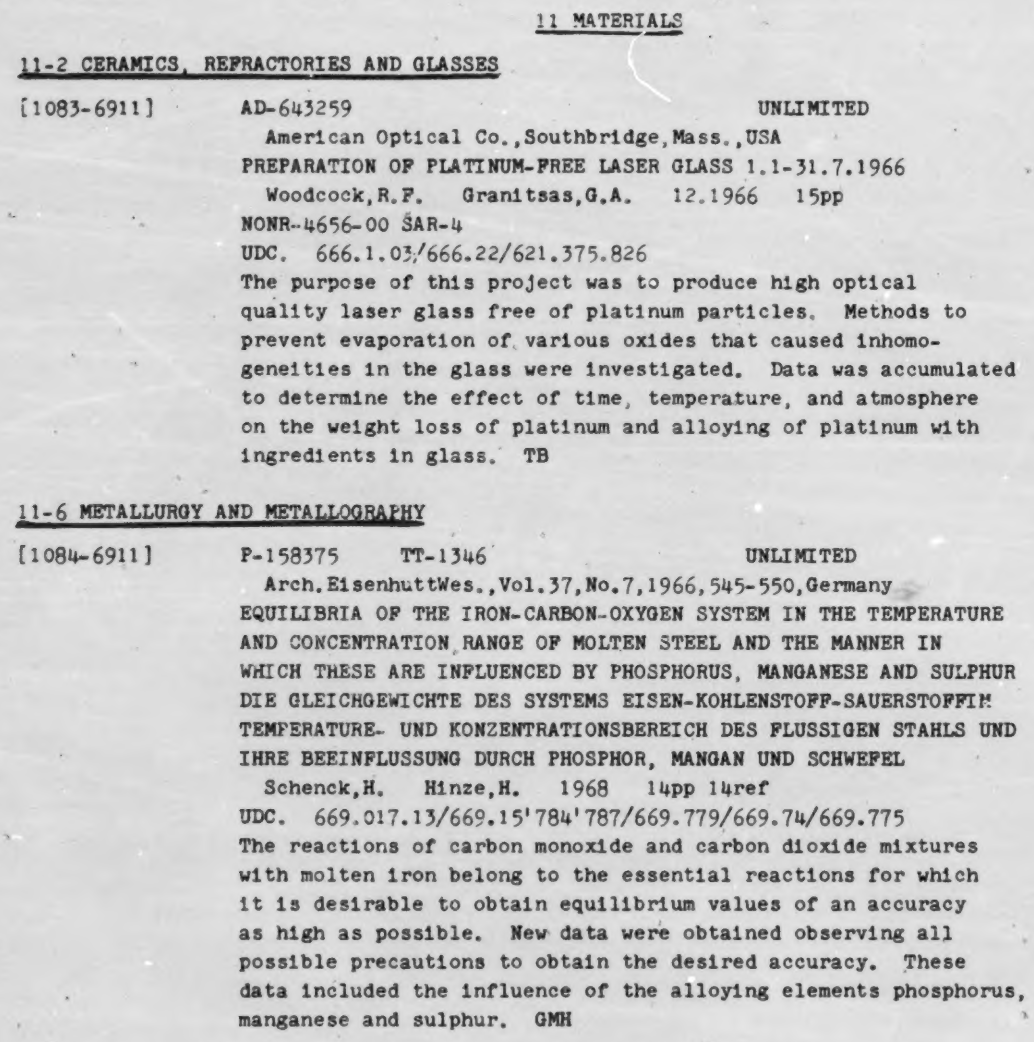
Many of the mechanical devices used in the creation and processing of research reports, announcement journals and similar documents can be described as office machinery. This includes addressing and enveloping machines, some kinds of printing machines, special cabinets for storing and displaying index cards and printed indexes, and photographic devices for creating pages of text for use as printing masters (plate-makers). Such equipment can be adapted to reduce, for instance, the tedium of information retrieval from files of index cards. Some of these card-based devices involve the provision of coloured tags or notched edges to indicate broad subject groups, and other more sophisticated processes rely upon the coincidence of holes drilled, with considerable precision, in special cards. The computer can handle this kind of basic information in a much more versatile way, and rapid developments have taken place in the field since the first files of bibliographic information were set up in the computer by reading in data from punched cards and punched paper tape.

2.1 Tape Typewriters

It was the tape typewriter which pointed the way towards the computer-based scientific and technical information (STI) system and operations of this kind are described in detail in Section 4 of this Manual. The simultaneous creation of a set of data in the form of text ('hardcopy') and as machine-readable codes punched onto paper tape, not only eliminated repetitive typing but provided a record which was acceptable to the computer. Control codes related to a second punch, for instance, can be used to create a special paper tape containing selected fields which, when sorted by means of standard software, provide the basis of effective and economical printed indexes. The judicious selection of a clear type-face and a well-designed format control program loaded into an auxiliary reader can ensure a satisfactory page layout for the main text of an announcement journal (see Figure 1), and index pages created automatically and printed by the computer have proved acceptable even when photo-reduced to conform to the required page size. The implications of the photo-reduction of such pages are dealt with in detail in paragraph 4.18.2, but it can be noted here that, where facilities exist for the photo-reduction of text according to a variable scale, a two-column ('two-up') page format can readily be achieved by the appropriate arrangement of the final text output from the typewriter at the camera copying stage. Useful title strips can also be added at this stage to provide extra subject headings which will link the text with the index (see Figure 2).

2.2 Computerised Indexes via the Tape Typewriter

"Per issue" indexes of the kind referred to above are generally printed from transient files of information held in the computer, that is, files which will be overwritten by the next increment of bibliographic data. Magnetic tape storage is relatively inexpensive, however, and provided that the computer has sufficient sorting capacity, the tapes can subsequently be used very economically to produce cumulated semi-annual, annual or volume indexes by running the contents of several of them together. These indexes can be photo-reduced to a convenient page size (generally A4 in the UK) and bound to form a separate publication which is useful for information retrieval purposes since it eliminates the need for manual searching through incremental indexes. A page from a typical index is shown in Figure 3.



11 MATERIALS

11-2 CERAMICS, REFRACTORIES AND GLASSES

[1083-6911] AD-643259 UNLIMITED
American Optical Co.,Southbridge,Mass.,USA
PREPARATION OF PLATINUM-FREE LASER GLASS 1.1-31.7.1966
Woodcock,R.F. Granitsas,G.A. 12.1966 15pp
NONR-4656-00 SAR-4
UDC. 666.1.03/666.22/621.375.826

The purpose of this project was to produce high optical quality laser glass free of platinum particles. Methods to prevent evaporation of various oxides that caused inhomogeneities in the glass were investigated. Data was accumulated to determine the effect of time, temperature, and atmosphere on the weight loss of platinum and alloying of platinum with ingredients in glass. TB

11-6 METALLURGY AND METALLOGRAPHY

[1084-6911] P-158375 TT-1346 UNLIMITED
Arch.EisenhuttWes.,Vol.37,No.7,1966,545-550,Germany
EQUILIBRIA OF THE IRON-CARBON-OXYGEN SYSTEM IN THE TEMPERATURE AND CONCENTRATION RANGE OF MOLTEN STEEL AND THE MANNER IN WHICH THESE ARE INFLUENCED BY PHOSPHORUS, MANGANESE AND SULPHUR
DIE GLEICHGEWICHTE DES SYSTEMS EISEN-KOHLENSTOFF-SAUERSTOFFIM TEMPERATURE- UND KONZENTRATIONSBEREICH DES FLUSSIGEN STAHLS UND IHRE BEEINFLUSSUNG DURCH PHOSPHOR, MANGAN UND SCHWEFEL
Schenck,H. Hinze,H. 1968 14pp 14ref
UDC. 669.017.13/669.15'784'787/669.779/669.74/669.775

The reactions of carbon monoxide and carbon dioxide mixtures with molten iron belong to the essential reactions for which it is desirable to obtain equilibrium values of an accuracy as high as possible. New data were obtained observing all possible precautions to obtain the desired accuracy. These data included the influence of the alloying elements phosphorus, manganese and sulphur. GMH

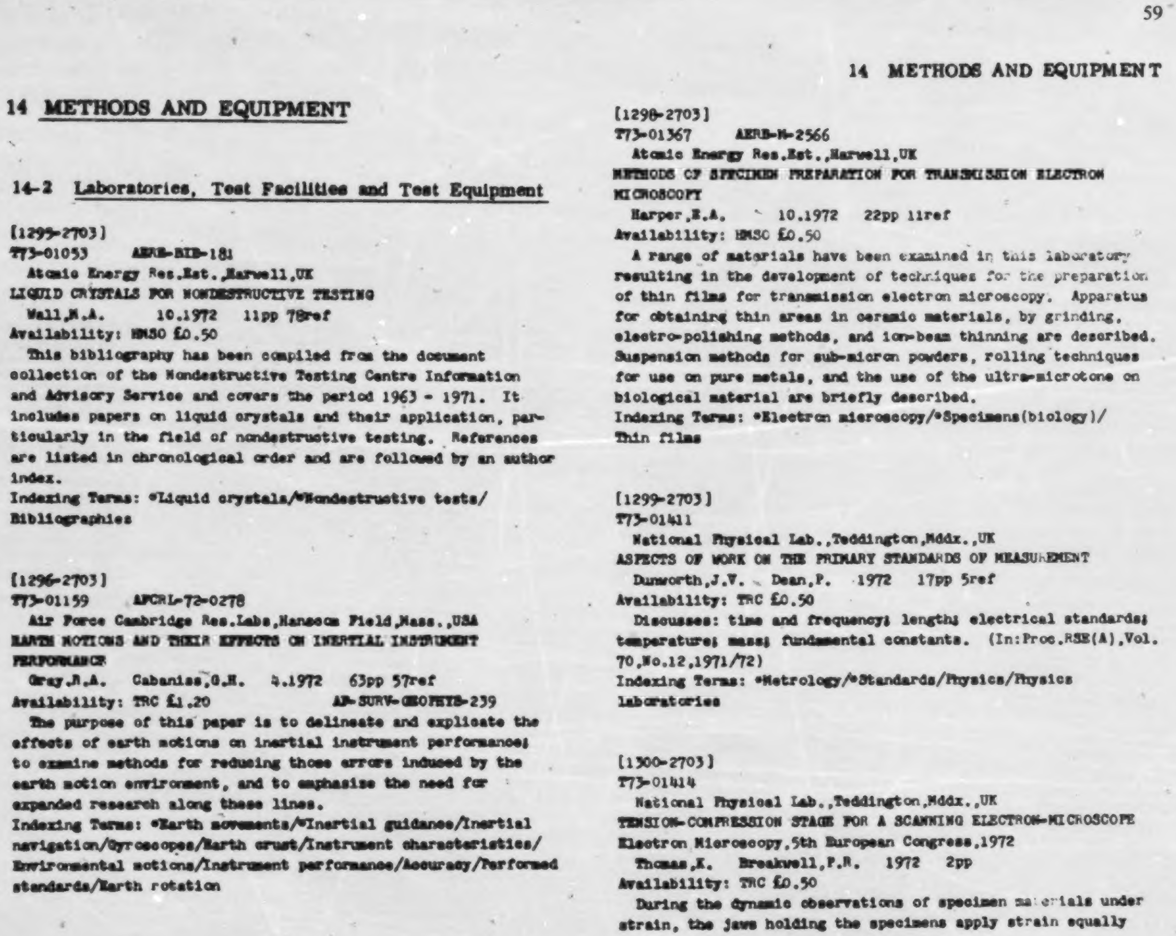
Fig.1 Single-column page of text from abstract journal
[tape-typewriter] – TRC 1969

2.3 Computer-based Data Stores

The storage of textual information in machine-readable (digital) form is becoming less expensive as techniques are developed for the economical use of large disc stores. Although the computer-held files used specifically for the production of printed indexes to announcement journals are transient the original files containing complete bibliographic references become a valuable resource when held in suitable computer stores, and most of the data bases in common use began in this way.

3. THE DATA BASE

The computer-held data base of bibliographic information is essentially a large dyanmic file, made dynamic by the frequent addition of new data, and in some cases, by the removal of redundant material. Figure 4 provides an indication of the way in which the file is built up, and the Data Conversion stage is included to emphasize the necessity for adopting a standard data format. When a set of elements of bibliographic information is put together to form a record, or item, in a data base file, each element becomes a field in that record and must be input according to strict rules in order to conform to the standard format. Data format rules are referred to in paragraphs 4.6 and 4.7, and a typical data base file format is similar to the main bibliographic record shown in Figure 5. Once a data base format has been established it can be very costly to change it in order to include additional elements, but conversion from one format, containing the basic bibliographic data fields in a specific sequence, to another in which the same elements appear in a somewhat different form or different order (eg the standard data base format adopted for the system), is relatively straightforward in a computer context. Thus a single data base can be made up of information derived in machine-readable form from



59

14 METHODS AND EQUIPMENT

14 METHODS AND EQUIPMENT

14-2 Laboratories, Test Facilities and Test Equipment

[1295-2703]
T73-01053   AERE-BIB-181
Atomic Energy Res.Est.,Harwell,UK
LIQUID CRYSTALS FOR NONDESTRUCTIVE TESTING
Wall,M.A.   10.1972   11pp 78ref
Availability: HMSO £0.50
This bibliography has been compiled from the document collection of the Nondestructive Testing Centre Information and Advisory Service and covers the period 1963 - 1971. It includes papers on liquid crystals and their application, particularly in the field of nondestructive testing. References are listed in chronological order and are followed by an author index.
Indexing Terms: *Liquid crystals/*Nondestructive tests/Bibliographies

[1296-2703]
T73-01159   AFCRL-72-0278
Air Force Cambridge Res.Labs,Hanscom Field,Mass.,USA
EARTH MOTIONS AND THEIR EFFECTS ON INERTIAL INSTRUMENT PERFORMANCE
Gray,R.A.   Cabaniss,G.H.   4.1972   63pp 57ref
Availability: TRC £1.20   AF-SURV-GEOPHYS-239
The purpose of this paper is to delineate and explicate the effects of earth motions on inertial instrument performance; to examine methods for reducing those errors induced by the earth motion environment, and to emphasize the need for expanded research along these lines.
Indexing Terms: *Earth movements/*Inertial guidance/Inertial navigation/Gyroscopes/Earth crust/Instrument characteristics/Environmental motions/Instrument performance/Accuracy/Performed standards/Earth rotation

[1298-2703]
T73-01367   AERE-M-2566
Atomic Energy Res.Est.,Harwell,UK
METHODS OF SPECIMEN PREPARATION FOR TRANSMISSION ELECTRON MICROSCOPY
Harper,E.A.   10.1972   22pp 11ref
Availability: HMSO £0.50
A range of materials have been examined in this laboratory resulting in the development of techniques for the preparation of thin films for transmission electron microscopy. Apparatus for obtaining thin areas in ceramic materials, by grinding, electro-polishing methods, and ion-beam thinning are described. Suspension methods for sub-micron powders, rolling techniques for use on pure metals, and the use of the ultra-microtone on biological material are briefly described.
Indexing Terms: *Electron microscopy/*Specimens(biology)/Thin films

[1299-2703]
T73-01411
National Physical Lab.,Teddington,Mddx.,UK
ASPECTS OF WORK ON THE PRIMARY STANDARDS OF MEASUREMENT
Dunworth,J.V.   Dean,P.   1972   17pp 5ref
Availability: TRC £0.50
Discusses: time and frequency; length; electrical standards; temperature; mass; fundamental constants. (In:Proc.RSE(A),Vol.70,No.12,1971/72)
Indexing Terms: *Metrology/*Standards/Physics/Physics laboratories

[1300-2703]
T73-01414
National Physical Lab.,Teddington,Mddx.,UK
TENSION-COMPRESSION STAGE FOR A SCANNING ELECTRON-MICROSCOPE
Electron Microscopy,5th European Congress,1972
Thomas,K.   Breakwell,P.R.   1972   2pp
Availability: TRC £0.50
During the dynamic observations of specimen materials under strain, the jaws holding the specimens apply strain equally

Two-single-column pages were laid side-by-side under the camera and photo-reduced to 75% of the original size.

Heading and sub-heading strips were added as required before each page was photographed.

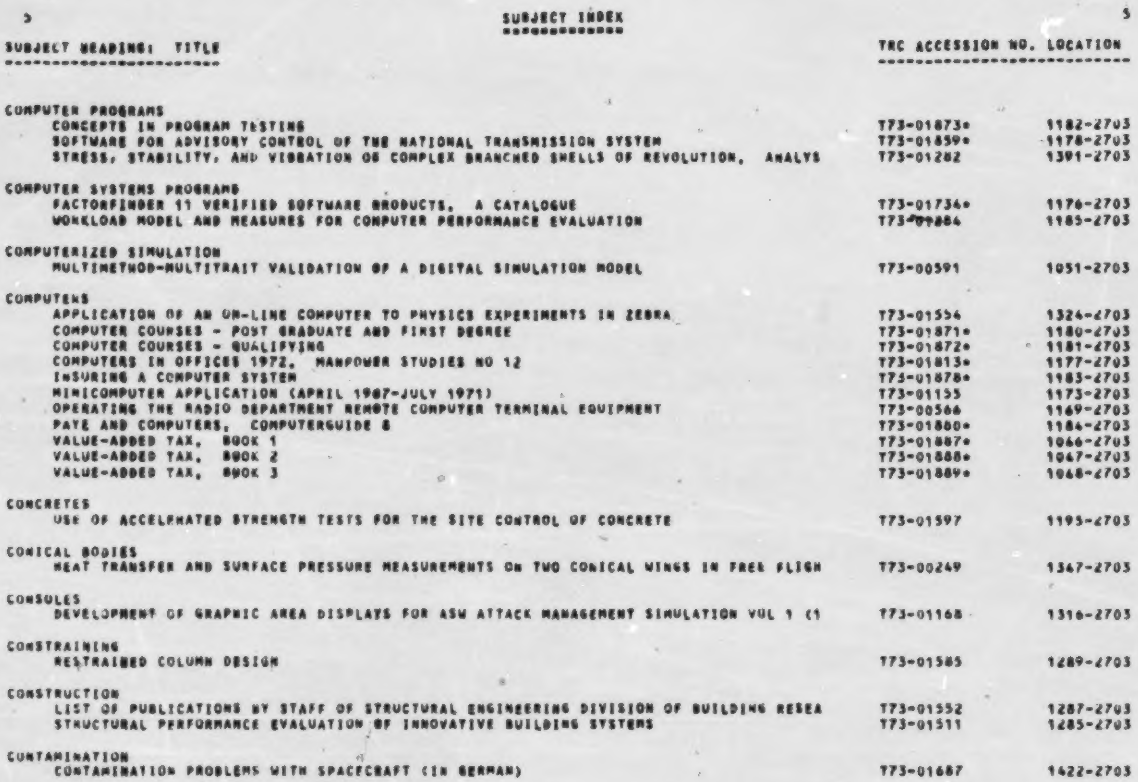
Fig.2 Double-column page of text from abstract journal
[tape-typewriter] – TRC 1973

a variety of sources. Attempts continue to be made to encourage data base producers, particularly those working in the fields of science and technology, to adopt a standard format at source, but experience indicates that such principles tend to become ponderous and inflexible in practice. The MARC (MAchine Readable Cataloguing) system was developed by the US Library of Congress and the British Library to provide for the international exchange of bibliographic information and this is probably the most notable example of a standard data format.

Although large disc stores can provide direct access to vast quantities of bibliographic information when held on-line, it is usual for the latest accessions to be in the greatest demand by information users. This leads to the concept of "levels of storage" which can be applied to effect useful economies. The latest information can be held on disc, with the older information which is less in demand being available on magnetic tape ready for transfer to disc in response to a specific request. Users of such systems are expected to define a range of dates when requesting information, and any charges made for transferring data from tape to disc and back again to tape are reflected in the overall costings.

3.1 Storage of Data Within the Computer (see Glossary of Terms)

When setting out to define the optimum size of an individual record within a disc file, it is essential to take account of the basic hardware design. An ICL 1900-series computer, for instance, uses a 24-bit 'word', a 128-word 'block', and 'buckets' containing 1, 2 4 or 8 blocks of data, and to achieve economy in the use of storage space within the machine or its peripherals it is advisable to set up files of such a size that they fit closely within these boundaries. Techniques have been developed for compressing the data in digital form within individual records to avoid waste of space, but this procedure is not so effective with text as it is with files of numeric data.



```
 >                                SUBJECT INDEX                                                           5
                                  *************
SUBJECT HEADING:  TITLE                                                         TRC ACCESSION NO. LOCATION
----------------------                                                          --------------------------

COMPUTER PROGRAMS
     CONCEPTS IN PROGRAM TESTING                                                T73-01873*       1182-2703
     SOFTWARE FOR ADVISORY CONTROL OF THE NATIONAL TRANSMISSION SYSTEM          T73-01859*       1178-2703
     STRESS, STABILITY, AND VIBRATION OF COMPLEX BRANCHED SHELLS OF REVOLUTION,  ANALYS   T73-01282   1391-2703

COMPUTER SYSTEMS PROGRAMS
     FACTORFINDER 11 VERIFIED SOFTWARE PRODUCTS,  A CATALOGUE                   T73-01736*       1176-2703
     WORKLOAD MODEL AND MEASURES FOR COMPUTER PERFORMANCE EVALUATION            T73-01884        1185-2703

COMPUTERIZED SIMULATION
     MULTIMETHOD-MULTITRAIT VALIDATION OF A DIGITAL SIMULATION MODEL            T73-00591        1051-2703

COMPUTERS
     APPLICATION OF AN ON-LINE COMPUTER TO PHYSICS EXPERIMENTS IN ZEBRA         T73-01556        1326-2703
     COMPUTER COURSES - POST GRADUATE AND FIRST DEGREE                          T73-01871*       1180-2703
     COMPUTER COURSES - QUALIFYING                                              T73-01872*       1181-2703
     COMPUTERS IN OFFICES 1972,  MANPOWER STUDIES NO 12                         T73-01813*       1177-2703
     INSURING A COMPUTER SYSTEM                                                 T73-01878*       1183-2703
     MINICOMPUTER APPLICATION (APRIL 1967-JULY 1971)                            T73-01155        1173-2703
     OPERATING THE RADIO DEPARTMENT REMOTE COMPUTER TERMINAL EQUIPMENT          T73-00566        1169-2703
     PAYE AND COMPUTERS,  COMPUTERGUIDE 8                                       T73-01880*       1184-2703
     VALUE-ADDED TAX,  BOOK 1                                                   T73-01887*       1066-2703
     VALUE-ADDED TAX,  BOOK 2                                                   T73-01888*       1067-2703
     VALUE-ADDED TAX,  BOOK 3                                                   T73-01889*       1068-2703

CONCRETES
     USE OF ACCELERATED STRENGTH TESTS FOR THE SITE CONTROL OF CONCRETE         T73-01597        1195-2703

CONICAL BODIES
     HEAT TRANSFER AND SURFACE PRESSURE MEASUREMENTS ON TWO CONICAL WINGS IN FREE FLIGH   T73-00249   1367-2703

CONSOLES
     DEVELOPMENT OF GRAPHIC AREA DISPLAYS FOR ASW ATTACK MANAGEMENT SIMULATION VOL 1 (1   T73-01168   1316-2703

CONSTRAINING
     RESTRAINED COLUMN DESIGN                                                   T73-01585        1289-2703

CONSTRUCTION
     LIST OF PUBLICATIONS BY STAFF OF STRUCTURAL ENGINEERING DIVISION OF BUILDING RESEA   T73-01552   1287-2703
     STRUCTURAL PERFORMANCE EVALUATION OF INNOVATIVE BUILDING SYSTEMS           T73-01511        1285-2703

CONTAMINATION
     CONTAMINATION PROBLEMS WITH SPACECRAFT (IN GERMAN)                         T73-01687        1422-2703
```

The indexes to each issue of the abstract journal were printed in upper case on an ICL 1900-series computer and photo-reduced to about 60% of the original size to fit conveniently on the page.

The input to the computer was the punched paper tape created by the tape-typewriter as the main text of the journal was being typed.

Fig.3 Extract from subject index – TRC 1973

Ideally, the abstract of each report held in the data base should form part of the record, but this can increase the record size from, say 400 characters to more than 1500 characters for the normal item of bibliographic information held in an STI system. Thus it is considerations of field size, record size, and file size which exercise a fundamental influence upon the practical and financial feasibility of any data base.

It has been pointed out that STI data bases are mostly built up of regular increments of data. This is quite a convenient arrangement, since it keeps the latest information, which is generally most in demand, in the most accessible position in the total file. Holding a data base in chronological order calls for reliable machine-held indexes to facilitate access by subject, author, report number, etc., and this makes the 'weeding-out' of individual report references a very complex matter.

3.2 The Computerised Thesaurus

Computerised information retrieval systems based upon a dictionary of approved indexing terms (descriptors) require a machine-readable thesaurus to facilitate subject searches. Figure 6 shows an extract from the EJC, Thesaurus of Engineering & Scientific Terms (TEST), and unlike the data base itself, the thesaurus file requires periodic reorganisation, with terms of various kinds being added, amended or deleted. It is convenient to control these activities by holding the master file on punched chards, and Figure 7 indicates the procedure for updating such a file with software designed to ensure that additional terms are posted correctly, i.e. are added not only as new main terms, but also as broader or narrower terms, suitably annotated, under their relevant main term headings. Considerable intellectual intervention is required to ensure that each new term is fully integrated into the thesaurus and a system of codes, to be punched into the cards in addition to the text, can be used to enable the update program to make the amendments as required. The handling of alterations to and deletions from the terms in the file is generally simpler than making additions, but precise planning is required to keep the thesaurus file up-to-date, and a complete update run is necessary at regular intervals when a check is also made to ensure that the latest set of interpreted punched cards is correct. On-line updating from a VDU terminal is feasible, but it is generally less reliable because of the greater possibility of human error.

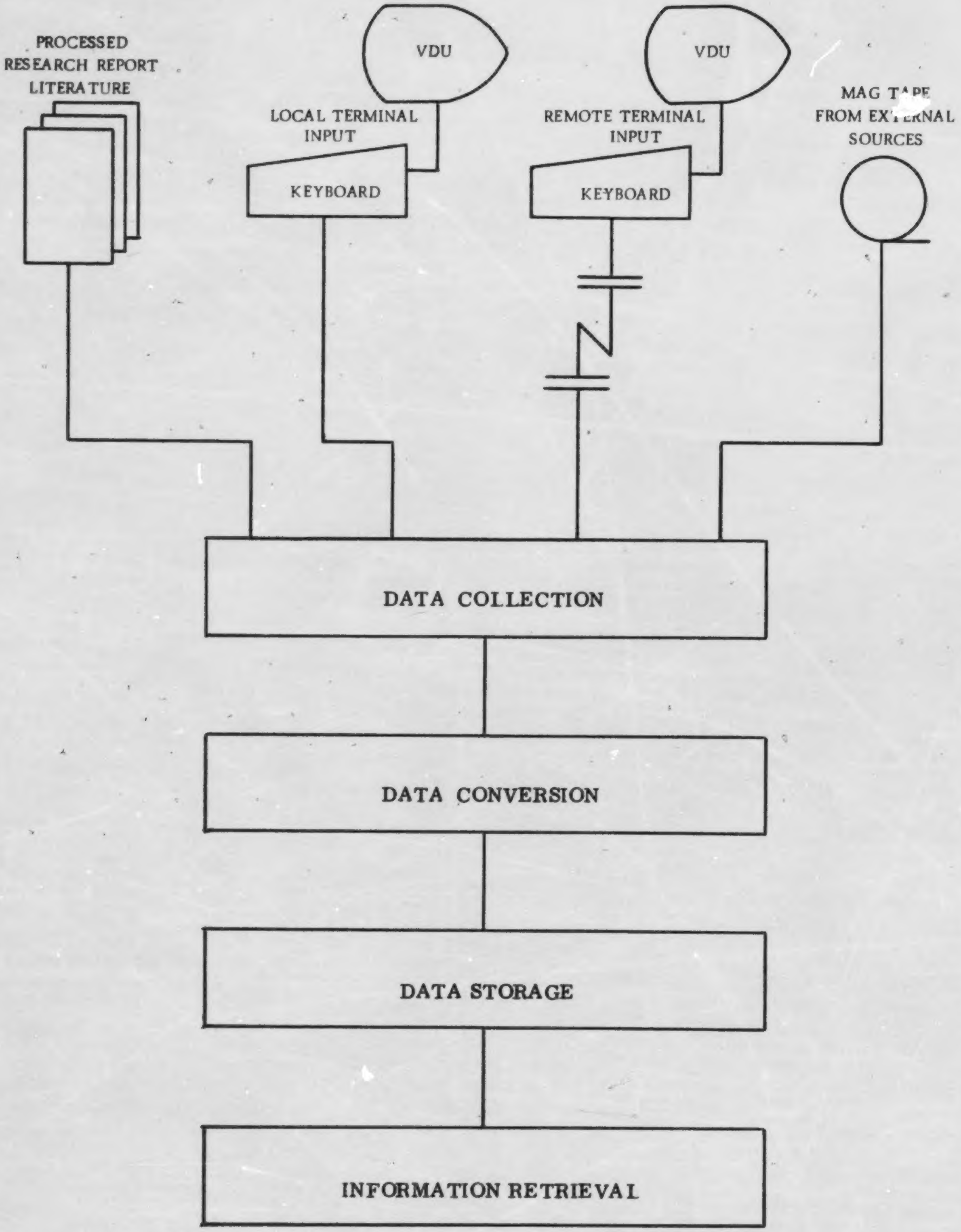

PROCESSED
RESEARCH REPORT
LITERATURE
VDU
LOCAL TERMINAL
INPUT
KEYBOARD
VDU
REMOTE TERMINAL
INPUT
KEYBOARD
MAG TAPE
FROM EXTERNAL
SOURCES
DATA COLLECTION
DATA CONVERSION
DATA STORAGE
INFORMATION RETRIEVAL

Fig.4 Database structure

Field Number	*Field Size (Characters)*	*Contents*
1	4	Main Subject Code
2	30	Accession Number
3	25	Originator's Reference Number
4	25	Agency Reference Number
5	5	Increment Number
6	150	Originator
7	300	Title in English
8	300	Title in Original Language
9	300	Conference Details
10	25	First Author
11	25	Second Author
12	9	Date of Report
13	17	Pages and References
14	29	Contract Number and Period Reference
15	25	Project Number
16	25	Other Reference Number
17	25	Availability and Price
18	25	Further Reference Number
19	2091	Abstract
20	565	Indexing Terms
	4000	

Notes:

1. Maximum record length = 4000 characters
   Average record length = 1500 characters
2. Each field size includes the end-of-field marker.
3. The field size quoted for Field 20 includes all sub-field markers and the end-of-record marker.
4. If the document is a translation the title field (Field 7) will end with the identity of the country of origin abbreviated as necessary and held between brackets preceded by a sub-field marker.

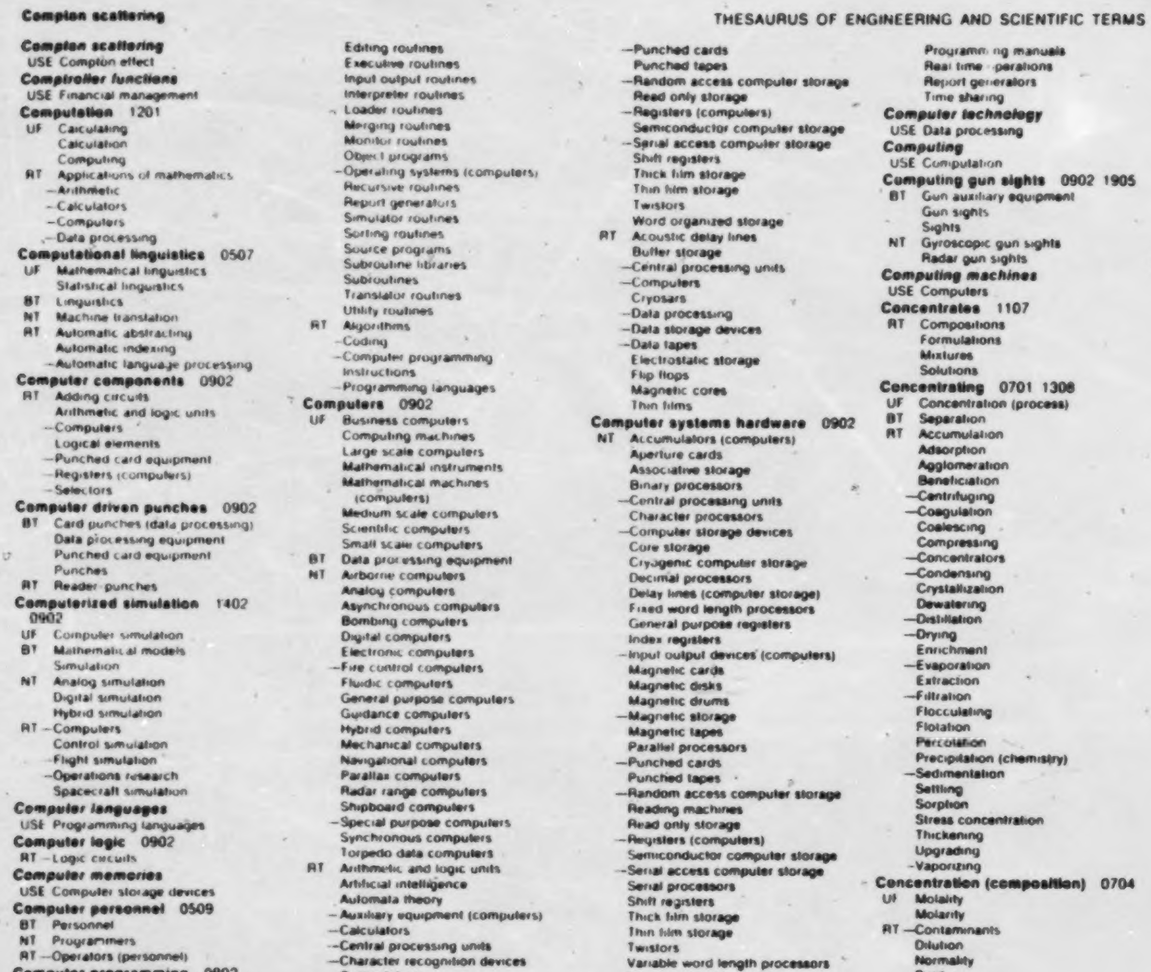
Fig.5 Main bibliographic record (input file structure)

4. THE IN-HOUSE COMPUTER

4.1 Initial Preparations

4.1.1 The Feasibility Study is the essential pre-requisite to any computer-based STI system. Depending upon the size of the installation and the time available, the study should be carried out by an individual or a small team with experience of the organisation and management of office procedures but with a greater knowledge of the implications of, and opportunities for, the automatic processing of textual data. Consultants from outside the organisation affected should be expected to confer with management at frequent intervals during the study to ensure that the plans develop in the right direction. In cases where the computer is being installed to take over or extend operations already being performed by manual methods or by some form of basic mechanisation, every step within each activity must be precisely documented and fully understood both as a separate function and also as it affects other operations within the total system. Careful analysis of this kind can show up ways in which current procedures may be improved, for it is often found that activities which are unnecessary, uneconomical or redundant have been allowed to become an accepted part of the system. Thus detailed systems analysis as a prelude to computerisation can prove useful even if the actual introduction of the computer is delayed.

4.1.2 Special Accommodation – Careful planning is needed to ensure that the special accommodation required to house the computer and its ancillary equipment is ready on time. A frequent source of delay is the creation of the controlled environment needed by the larger machines (generally those with more than 16 000 bytes of core), although many of the mini-computers referred to in Chapter five will operate satisfactorily in the kind of air-conditioning which is found in modern office accommodation. In most cases it will be necessary to build some special accommodation or to carry out modifications to an existing building, and it is wise to start on this part of the overall task at the earliest possible time and to monitor progress most carefully. As well as the machine itself there must be accommodation for computer operators, programmers and data preparation staff, all of whom must be recruited and trained in good time. The cost



Compton scattering

THESAURUS OF ENGINEERING AND SCIENTIFIC TERMS

Compton scattering
USE Compton effect
Comptroller functions
USE Financial management
Computation 1201
UF Calculating
Calculation
Computing
RT Applications of mathematics
—Arithmetic
—Calculators
—Computers
—Data processing
Computational linguistics 0507
UF Mathematical linguistics
Statistical linguistics
BT Linguistics
NT Machine translation
RT Automatic abstracting
Automatic indexing
—Automatic language processing
Computer components 0902
RT Adding circuits
Arithmetic and logic units
—Computers
Logical elements
—Punched card equipment
—Registers (computers)
—Selectors
Computer driven punches 0902
BT Card punches (data processing)
Data processing equipment
Punched card equipment
Punches
RT Reader punches
Computerized simulation 1402 0902
UF Computer simulation
BT Mathematical models
Simulation
NT Analog simulation
Digital simulation
Hybrid simulation
RT —Computers
Control simulation
—Flight simulation
—Operations research
Spacecraft simulation
Computer languages
USE Programming languages
Computer logic 0902
RT —Logic circuits
Computer memories
USE Computer storage devices
Computer personnel 0509
BT Personnel
NT Programmers
RT —Operators (personnel)
Computer programming 0902
UF Programming (computers)
Programming techniques
NT Control sequences
Machine coding
Macroprogramming
Microprogramming
Symbolic programming
RT Addressing
Algorithms
—Coding
—Computer programs
—Computer systems programs
—Data processing
Digital techniques
Dynamic programming
Flow charting
Flow charts
Heuristic methods
—Linear programming
Logic design
—Mathematical programming
Multiprocessing
Multiprogramming
—Numerical analysis
—Programming languages
Programming manuals
Real time operations
Sequencing
Systems analysis
Time sharing
Computer programs 0902
UF Programs (computers)
Routines (computer programming)
NT Assembler routines
Compilers
—Computer systems programs
Diagnostic routines
Editing routines
Executive routines
Input output routines
Interpreter routines
Loader routines
Merging routines
Monitor routines
Object programs
—Operating systems (computers)
Recursive routines
Report generators
Simulator routines
Sorting routines
Source programs
Subroutine libraries
Subroutines
Translator routines
Utility routines
RT Algorithms
—Coding
—Computer programming
Instructions
—Programming languages
Computers 0902
UF Business computers
Computing machines
Large scale computers
Mathematical instruments
Mathematical machines (computers)
Medium scale computers
Scientific computers
Small scale computers
BT Data processing equipment
NT Airborne computers
Analog computers
Asynchronous computers
Bombing computers
Digital computers
Electronic computers
—Fire control computers
Fluidic computers
General purpose computers
Guidance computers
Hybrid computers
Mechanical computers
Navigational computers
Parallax computers
Radar range computers
Shipboard computers
—Special purpose computers
Synchronous computers
Torpedo data computers
RT Arithmetic and logic units
Artificial intelligence
Automata theory
—Auxiliary equipment (computers)
—Calculators
—Central processing units
—Character recognition devices
Computation
Computer components
—Computerized simulation
—Computer storage devices
—Computer systems hardware
—Computer systems programs
Cybernetics
—Data processing
Information systems
—Logic circuits
Logic design
—Pattern recognition devices
Printers (data processing)
Punched tape readers
Systems management
—Telecommunication
Computer simulation
USE Computerized simulation
Computer software
USE Computer systems programs
Computer storage devices 0902
UF Computer memories
Memory devices
Memory systems
BT Computer systems hardware
NT Accumulators (computers)
Aperture cards
Associative storage
Core storage
Cryogenic computer storage
Delay lines (computer storage)
General purpose registers
Index registers
Magnetic cards
Magnetic disks
Magnetic drums
—Magnetic storage
Magnetic tapes
—Punched cards
Punched tapes
—Random access computer storage
Read only storage
—Registers (computers)
Semiconductor computer storage
—Serial access computer storage
Shift registers
Thick film storage
Thin film storage
Twistors
Word organized storage
RT Acoustic delay lines
Buffer storage
—Central processing units
—Computers
Cryosars
—Data processing
—Data storage devices
—Data tapes
Electrostatic storage
Flip flops
Magnetic cores
Thin films
Computer systems hardware 0902
NT Accumulators (computers)
Aperture cards
Associative storage
Binary processors
—Central processing units
Character processors
—Computer storage devices
Core storage
Cryogenic computer storage
Decimal processors
Delay lines (computer storage)
Fixed word length processors
General purpose registers
Index registers
—Input output devices (computers)
Magnetic cards
Magnetic disks
Magnetic drums
—Magnetic storage
Magnetic tapes
Parallel processors
—Punched cards
Punched tapes
—Random access computer storage
Reading machines
Read only storage
—Registers (computers)
Semiconductor computer storage
—Serial access computer storage
Serial processors
Shift registers
Thick film storage
Thin film storage
Twistors
Variable word length processors
Word organized storage
RT —Auxiliary equipment (computers)
—Computers
Consoles
Control circuits
Data links
—Data processing
—Data processing equipment
—Data storage devices
—Logic circuits
Magnetic tape transports
Multiprocessing
Multiprogramming
Real time operations
Time sharing
Computer systems programs 0902
UF Computer software
Software (computers)
BT Computer programs
NT Assembler routines
Compilers
Executive routines
Input output routines
Interpreter routines
Loader routines
Monitor routines
—Operating systems (computers)
Simulator routines
Subroutine libraries
Translator routines
RT —Computer programming
—Computers
—Data processing
Diagnostic routines
Editing routines
Multiprocessing
Multiprogramming
—Programming languages
Programming manuals
Real time operations
Report generators
Time sharing
Computer technology
USE Data processing
Computing
USE Computation
Computing gun sights 0902 1905
BT Gun auxiliary equipment
Gun sights
Sights
NT Gyroscopic gun sights
Radar gun sights
Computing machines
USE Computers
Concentrates 1107
RT Compositions
Formulations
Mixtures
Solutions
Concentrating 0701 1308
UF Concentration (process)
BT Separation
RT Accumulation
Adsorption
Agglomeration
Beneficiation
—Centrifuging
—Coagulation
Coalescing
Compressing
—Concentrators
—Condensing
Crystallization
Dewatering
—Distillation
—Drying
Enrichment
—Evaporation
Extraction
—Filtration
Flocculating
Flotation
Percolation
Precipitation (chemistry)
—Sedimentation
Settling
Sorption
Stress concentration
Thickening
Upgrading
—Vaporizing
Concentration (composition) 0704
UF Molality
Molarity
RT —Contaminants
Dilution
Normality
Purity
—Quality
—Sampling
Solubility
Solvent action
Concentration (process)
USE Concentrating
Concentration (stress)
USE Stress concentration
Concentration cell corrosion 1113
UF Crevice corrosion
BT Corrosion
Corrosion mechanisms
RT Corrosion fatigue
Erosion corrosion
Fretting corrosion
—Galvanic corrosion
Impingement corrosion
Intergranular corrosion
Stray current corrosion
Stress corrosion
Concentrators 0701 1309
BT Separators
NT —Centrifuges
Dense media separators
Dialyzers
—Electric concentrators
Electrostatic separators
Filter presses
Flotation machines
—Gravity concentrators
Hydrocyclones
Jigs (concentrators)
Magnetic separators
Settlers (separators)
Sluices (concentrators)
Solid bowl classifiers
Spirals (concentrators)

90

Subject Category Index numbers follow main terms. (—) = See main entry for narrower terms. ‡ = Consult main entry.

Fig.6 Page from EJC (TEST) thesaurus

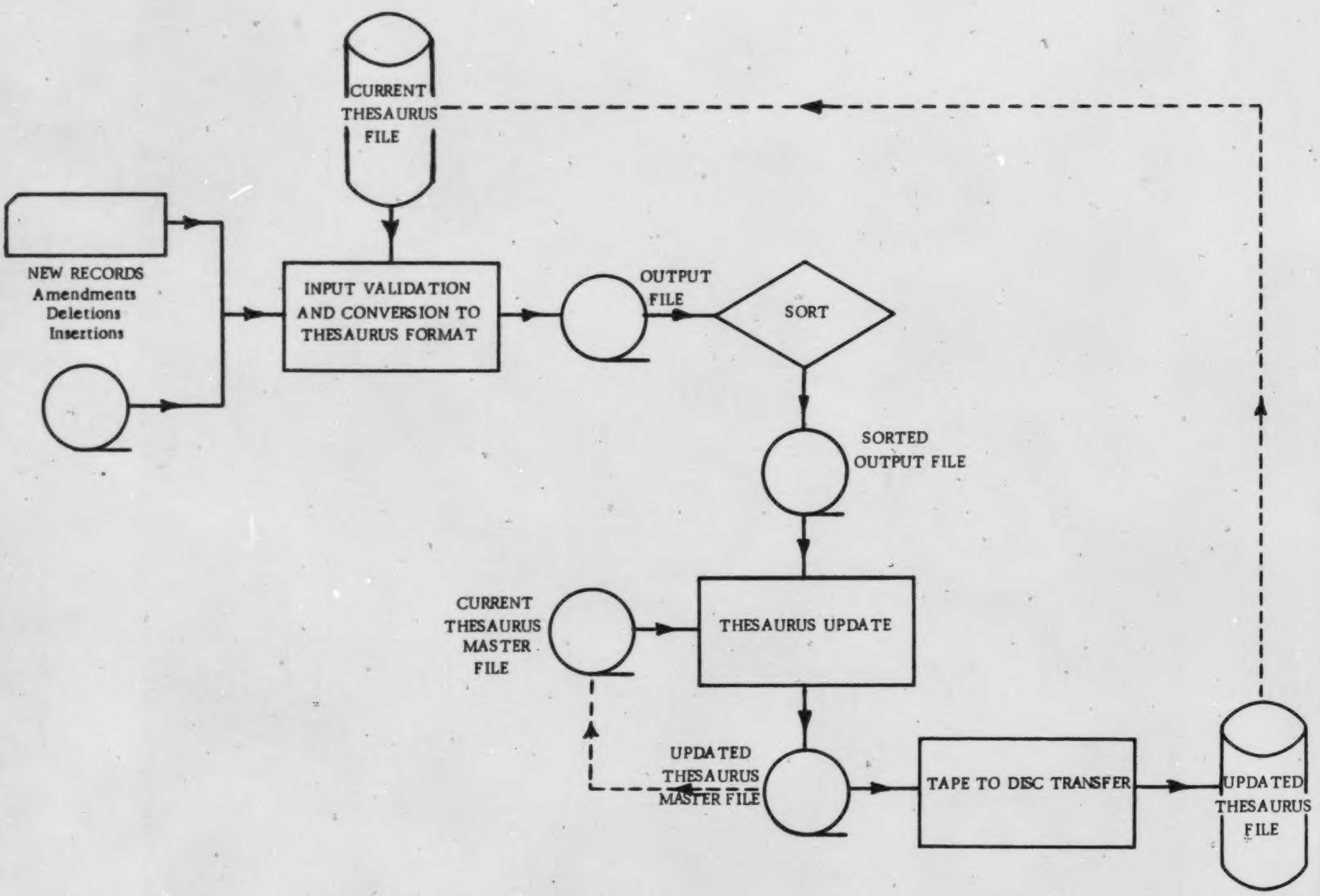

CURRENT
THESAURUS
FILE
NEW RECORDS
Amendments
Deletions
Insertions
INPUT VALIDATION AND CONVERSION TO THESAURUS FORMAT
OUTPUT FILE
SORT
SORTED OUTPUT FILE
CURRENT THESAURUS MASTER FILE
THESAURUS UPDATE
UPDATED THESAURUS MASTER FILE
TAPE TO DISC TRANSFER
UPDATED THESAURUS FILE

Fig.7 Thesaurus update

of introducing a computer is such that it might prove uneconomical to let the machine stand idle, so that the provision of a rest-room and a simple kitchen should be considered at an early stage in order to facilitate shift working.

4.1.3 Back-up facilities – Although the computer aspects of STI handling are highly specialised in many respects, means should be sought of carrying out as many operations as possible on an alternative machine in the event of a serious breakdown. Subject to security considerations, this objective can sometimes be achieved by setting up a system of mutual help between the STI organisation concerned and a research establishment or similar institution which has a computer of suitable size and of the same manufacture.

4.1.4 Internal reorganisation – When the main constraints of budget, scope and time-scale have been laid down for the computerised system, it is vital to set up procedures for regular communication between the newly formed Automatic Data Processing (ADP) section and the internal groups most affected by the inevitable function re-organisation which will take place. A functional chart which could apply to an Information Centre is shown in Figure 8, and it will be seen that liaison is necessary at all levels of management. This is particularly important where the new system is to be superimposed upon well-established procedures with the object of ultimately replacing them. If the first output from the computer can be shown to bring about an obvious improvement to an existing operation, or to create a new product which fulfils a long-felt need, then those members of staff whose daily work is to be affected by the machine will be more likely to take a sympathetic view of subsequent developments.

4.2 Input File Structures

It has already been pointed out that the computer storage of bibliographic records which include abstracts can create problems of capacity and cost. These problems become greater if an attempt is made to store the complete text of reports, but they are not altogether insoluble given adequate resources, and advanced systems of this kind will be referred to when considering the implications of information retrieval later in this chapter.

There are many types of file structure which can be used to hold bibliographic records in the computer, but there are two basic kinds which will be described in some detail. The ideal system will probably incorporate aspects of both and, as knowledge and expertise develop, refinements such as data compression techniques can be introduced.

4.2.1 The single bibliographic file – Provided that there are adequate resources, it is quite possible to store all the bibliographic information which is to be processed in the form of individual records in a single main bibliographic file. The research report literature processed by the Technology Reports Centre of the UK Department of Industry is handled in this way, and it is against this background that the more generalised systems described in this chapter are defined. A typical record is shown in Figure 5, and it can be seen that it consists of twenty fields, including the complete abstract, arranged in a manner which is dictated by the need to display the various elements in a convenient sequence on the VDU screen as the data are being input. The average length of such records is about 1500 characters, but the system should allow for records up to 4000 characters in length.

An important factor governing the sequence of the fields in this type of system is the maximum number of characters allowed in a single line of the display, remembering that each space is counted as a character. Many VDUs have a line length of 80 characters so that, for instance, the two author fields, the date and the pagination information, which together occupy a possible maximum of 79 characters inclusive of spaces will fit conveniently into a single line of text (see Figure 9). The software is therefore designed to accept these data and store them in that order. On the other hand, the originator field, known within many systems as the corporate author field, can contain up to 150 characters but rarely exceeds 80. The software will therefore handle as many lines as each field requires at the input stage and provide the correct display format when called upon to do so.

Experience within TRC has proved that there are no serious information retrieval problems created by restricting the number of authors identified in the input file to two. The field sizes throughout the record are rarely found to be inadequate, and it has been shown that, for quick reference purposes, most reports can be identified by the first thirty characters of the title.

Certain fields within the record will inevitably contain sub-fields, i.e. discrete elements within the field which must be flagged for use by the software for special purposes. The indexing terms (descriptor) field is constructed in such a way that the individual terms can be extracted (see paragraph 4.7.6) and listed for information retrieval purposes. Each descriptor is, therefore, a sub-field and is separated from its neighbour by a code, the oblique (/) in Figure 9 which is recognised by the software. In the example shown in Figure 10 the question mark (?) forms the field separator and the left-hand square bracket ([) is the record separator. The decision concerning the number of records which will constitute an incremental file will depend upon a wide range of factors, the most important of which are the average number of documents passing through the system in a given period (say a week), the average size of bibliographic record, the average number of records per printed page of the announcement journal, the required frequency of journal publication and the optimum number of pages per incremental issue of the journal. The data handling capacity of the computer and its peripherals will also be very significant, but the administrative decisions involved in these aspects of the total task of the STI system will have been taken into account in the course of the original feasibility study.

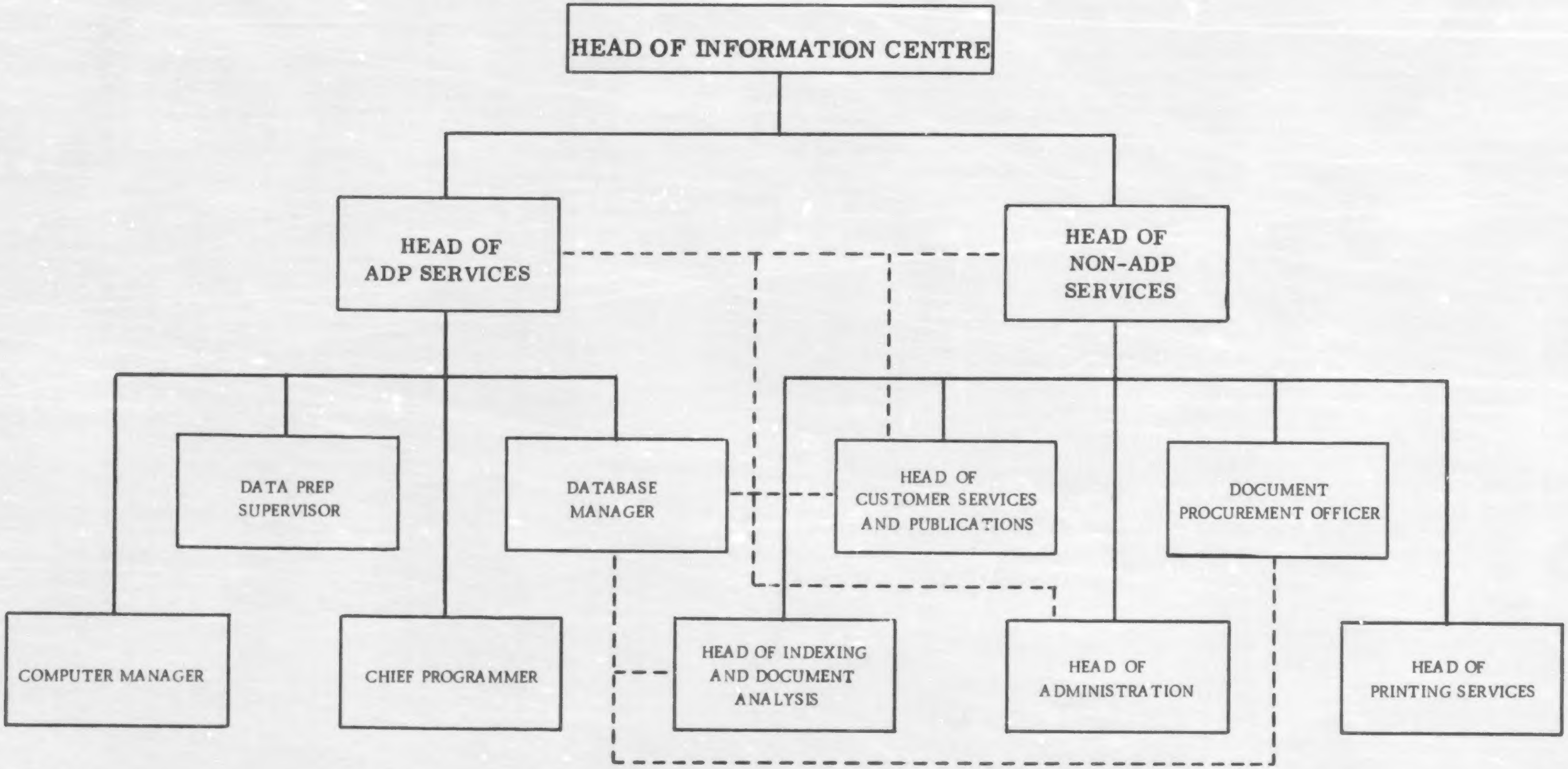
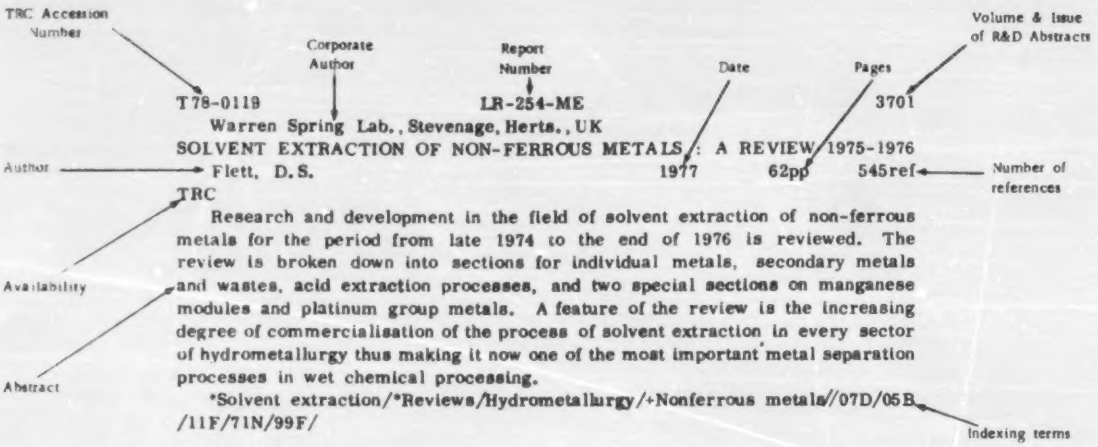

HEAD OF INFORMATION CENTRE
HEAD OF ADP SERVICES
HEAD OF NON-ADP SERVICES
DATA PREP SUPERVISOR
DATABASE MANAGER
HEAD OF CUSTOMER SERVICES AND PUBLICATIONS
DOCUMENT PROCUREMENT OFFICER
COMPUTER MANAGER
CHIEF PROGRAMMER
HEAD OF INDEXING AND DOCUMENT ANALYSIS
HEAD OF ADMINISTRATION
HEAD OF PRINTING SERVICES

Fig.8 A typical organisational structure (showing lines of communication)



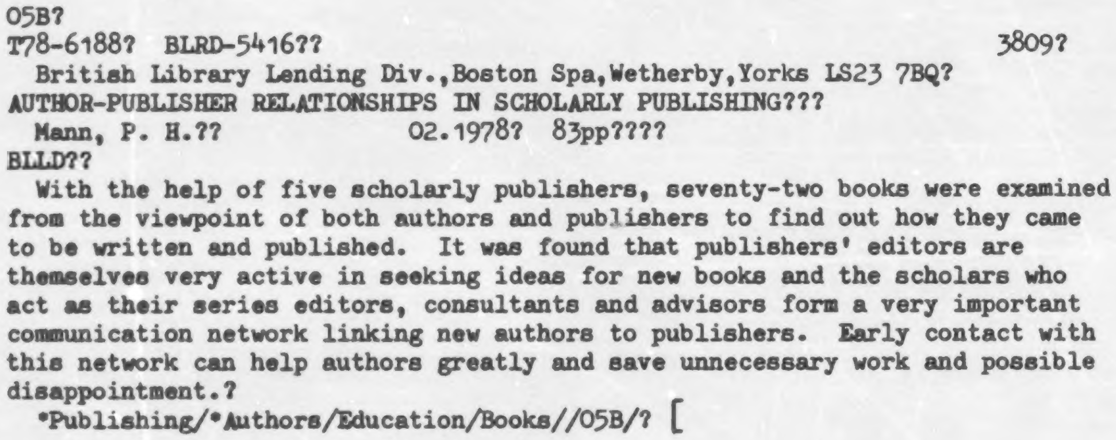
TRC Accession Number
Corporate Author
Report Number
Date
Pages
Volume & Issue of R&D Abstracts
Author
Number of references
Availability
Abstract
Indexing terms

```
T78-0119                           LR-254-ME                                3701
   Warren Spring Lab., Stevenage, Herts., UK
SOLVENT EXTRACTION OF NON-FERROUS METALS : A REVIEW 1975-1976
   Flett, D.S.                                     1977         62pp       545ref
TRC
   Research and development in the field of solvent extraction of non-ferrous
metals for the period from late 1974 to the end of 1976 is reviewed. The
review is broken down into sections for individual metals, secondary metals
and wastes, acid extraction processes, and two special sections on manganese
modules and platinum group metals. A feature of the review is the increasing
degree of commercialisation of the process of solvent extraction in every sector
of hydrometallurgy thus making it now one of the most important metal separation
processes in wet chemical processing.
   *Solvent extraction/*Reviews/Hydrometallurgy/+Nonferrous metals//07D/05B
/11F/71N/99F/
```

Indexing terms

- EJC-TEST thesaurus terms for subject index (*)
- other EJC-TEST terms
- non-EJC-TEST terms (+)
- Cosati code used for location in R & D ABSTRACTS (//)
- other Cosati and NTIS codes (/)

Fig.9 A typical abstract



```
05B?
T78-6188? BLRD-5416??                                                   3809?
  British Library Lending Div.,Boston Spa,Wetherby,Yorks LS23 7BQ?
AUTHOR-PUBLISHER RELATIONSHIPS IN SCHOLARLY PUBLISHING???
  Mann, P. H.??              02.1978? 83pp????
BLLD??
  With the help of five scholarly publishers, seventy-two books were examined
from the viewpoint of both authors and publishers to find out how they came
to be written and published. It was found that publishers' editors are
themselves very active in seeking ideas for new books and the scholars who
act as their series editors, consultants and advisors form a very important
communication network linking new authors to publishers. Early contact with
this network can help authors greatly and save unnecessary work and possible
disappointment.?
  *Publishing/*Authors/Education/Books//05B/? [
```

Fig.10 Bibliographic record displayed with control codes

Each incremental main bibliographic file is terminated by a unique code, say a double left hand square bracket ([[), and is handled as an entity in the system right through to the stage at which it is used for the printing of text and indexes. The data fields and control codes are then subjected to special software which converts the file to the formats required by the data base programs and the incremental file becomes part of the total data base system.

4.2.2 Parallel files – A useful alternative to the main bibliographic file is the system of parallel files. The simplest application of this system is to hold all fields except the abstract in one file (the Reports File) and to hold abstracts only, identified by the in-house document accession number, in a parallel file (the Abstracts File). The Reports File will be quite compact with each record containing 200 to 400 characters in the system shown in Figure 11, and it will have many uses within the total system, particularly for 'housekeeping' and other management purposes, e.g. the control of

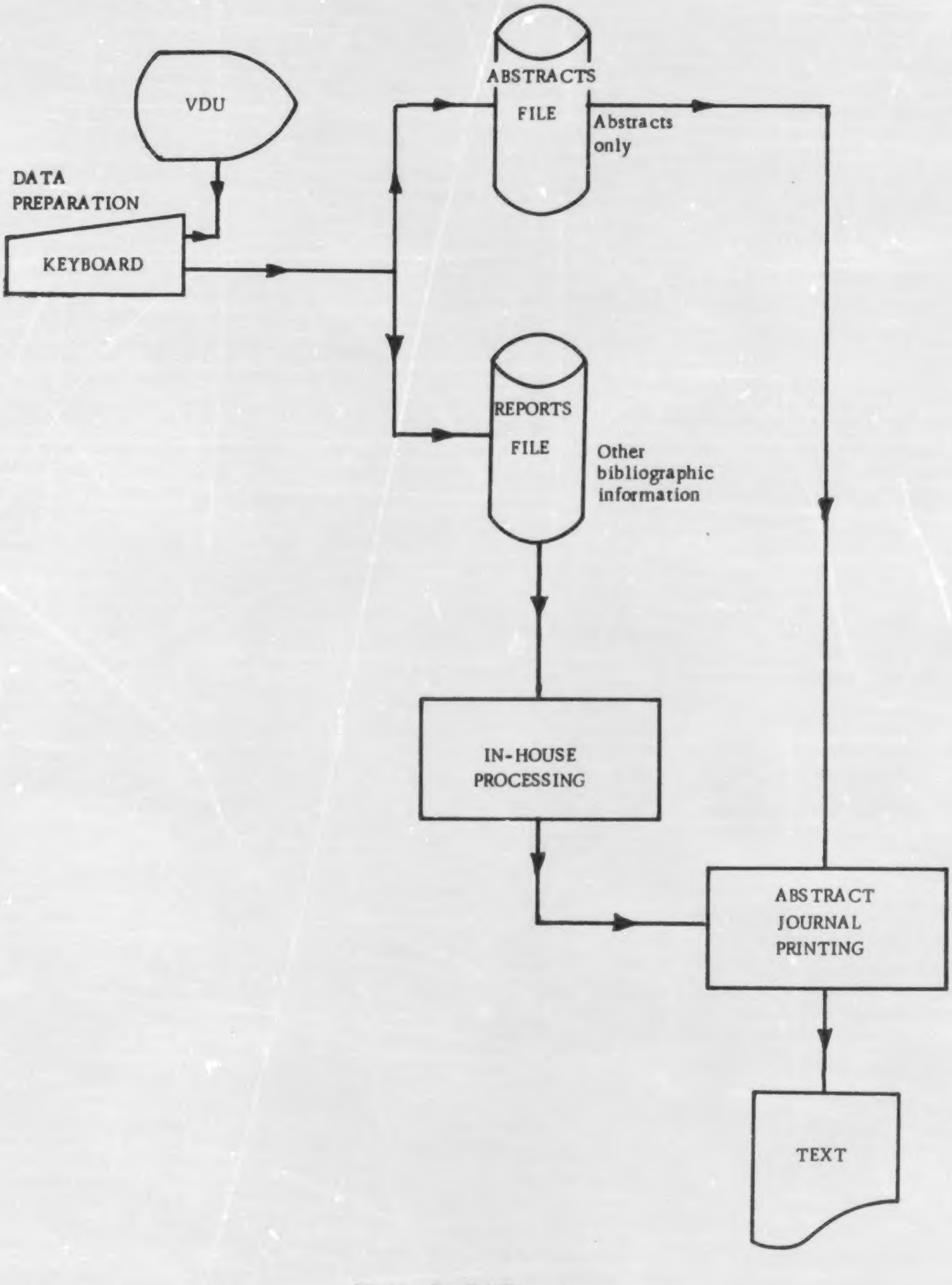

VDU
DATA
PREPARATION
KEYBOARD
ABSTRACTS
FILE
Abstracts
only
REPORTS
FILE
Other
bibliographic
information
IN-HOUSE
PROCESSING
ABSTRACT
JOURNAL
PRINTING
TEXT

Fig.11 Parallel file system

document issues and periodic statistics on work progress. Printing and display programs which call for abstracts will need to have the capability of linking in with the Abstracts File when complete bibliographic records are required.

The need for absolute precision on the part of the keyboard operators who create the input file is particularly significant in the case of those fields which hold reference number information, since a mis-quoted number can make a document irretrievable within the system. For systems employing punched cards or punched paper tape as input it is usual to key the data twice using different operators and then to compare the two sets of codes automatically.

The systems being described here, however, are based upon direct input from VDU terminals with single keying; it is then feasible to write alpha-numeric and character counting checks into validation software to protect the input against errors in fields which are susceptible to inaccuracies of this kind. Those fields which contain textual data can only be properly corrected by proof reading and text editing. Fields which form the main entries to indexes are particularly sensitive to keying errors, and this is especially true of the Corporate Author index based upon the Originator field. The slightest textual error, such as misquoted abbreviations or the use of incorrect punctuation will, if undetected, cause the computer to create a new heading identical in all other respects with the proper heading, thus impairing the accuracy and reliability of the index. There is a strong case, therefore, for establishing a master file of originators as an authority file cross-referenced to the file of bibliographic information, so that the keyboard operator needs only to insert a reference number in the originator field of the input file, leaving the software to pull out the corresponding corporate author information from the master file when a complete record is required.

A six-figure number, representing up to 999 999 different corporate authors, is considered sufficient for indexing purposes in the parallel file system, and it is far easier to apply a visual check to such a number than it is to proof-read the complete corporate author identity with its array of capital letters, special punctuation and abbreviations.

The system of parallel files could be extended to cover personal authors also, but such developments can generally be regarded as refinements.

4.3 Storage File Structures

The software which is used to count the data fields and to validate such fields as the accession number, main subject code and date in the single main bibliographic file system can be extended to re-sort the data elements when they have been corrected and to eliminate unwanted spaces from the record. Figure 12 shows a typical storage file, arranged with the main sort fields (main subject code, accession number, originator's reference number and related document numbers) at the top of each record. In most cases the first operation involving the storage file is the printing out of an increment of the semi-monthly announcement journal, and a sequential number (record number) is generated by the software for use as a control within the suite of programs which creates the printed output for each increment. The record number is not normally displayed.

Systems which employ parallel files for the storage of data can only operate efficiently if the data elements which make up the records are enhanced by the addition of special control codes (keys) and links at the data preparation stage. The software which controls the files will then recognise the keys and links and use them as pointers to the data elements in the various files when it is necessary to assemble records and display them in specific formats.

4.4 Output File Structure (Fig.13)

The principle function of the Output File in STI systems is to display bibliographic information either on the VDU screen or the printed page. The structure of the VDU displays will be described in paragraph 4.5 (Data Preparation), and the structure of the print file used to create the pages of a typical announcement journal will be referred to in paragraph 4.18 on printing. A variety of other output formats is required for data validation purposes and to provide brief details of the documents passing through the system for various housekeeping functions, so that it is important to ensure that control codes are retained within the text to provide keys for data selection and display. Programs can then be provided to list any individual fields or combination of fields for work-flow control purposes, e.g. from accession numbers only to the main bibliographic record complete with the symbols which represent the control codes. This latter feature is particularly important in a system which prepares output on magnetic or punched paper tapes for photo-typesetters.

4.5 Data Preparation Systems – General

Reference has already been made to the use of punched cards and punched paper tape as input to STI systems, and there are specific areas in which these media can be employed to advantage. The modern in-house computer generally lends itself to input via visual display units however, and there are great benefits to be gained from the instant display of text on the screen as the data are being captured in machine-readable form. The same broad rules governing the structure of the data elements within the record and the structure of the records within the file will apply whatever data preparation technique is employed, but the trend is towards entering data direct from VDU terminals, and it is the essential features of this type of system which will be described in detail.

Field Number	*Maximum Field Size (Characters)*	*Contents*
0	4	Record Count (Total number of characters in record)
1A	4	Record Number
1	4	Main Subject Code
2	30	Accession Number
3	25	Originator's Reference Number
4	25	Agency Reference Number
5	25	Project Number
6	25	Other Reference Number
7	25	Further Reference Number
8	29	Contract Number and Period Reference
9	9	Date of Report
10	17	Pages and References
11	25	First Author
12	25	Second Author
13	25	Availability and Price
14	5	Increment Number
15	150	Originator
16	300	Title in English
17	300	Title in Original Language
18	300	Conference Details
19	2091	Abstract
20	565	Indexing Terms
20A	1	Record Terminator

Fig.12 Typical storage file structure

Field Number	*Contents*
0	Record Count (Total number of characters in record)
1	Main Subject Code
2	Accession Number
3	Originator's Reference Number
4	Agency Reference Number
5	Project Number
6	Other Reference Number
7	Further Reference Number
8	Contract Number and Period Reference
9	Date of Report
10	Pages and References
11	First Author
12	Second Author
13	Availability and Price
14	Increment Number
15	Originator
16	Title in English
17	Title in Original Language
18	Conference Details
19	Abstract
20	Indexing Terms
20A	Record Terminator

Fig.13 Output file structure

4.6 Key-to-disc Input from Terminals Linked to an In-house Computer

A typical key-to-disc data prepration system is illustrated in Figures 14 and 15.The "Data Prep Room" in which the input VDU terminals are housed should be carpeted, quiet and comfortable to provide the operators with working conditions in which they can carry out their very exacting work with the minimum of distraction. It is particularly important to arrange the lighting in such a way that eye-strain is eliminated. Ideally the Data Prep Room should be near enough to

74

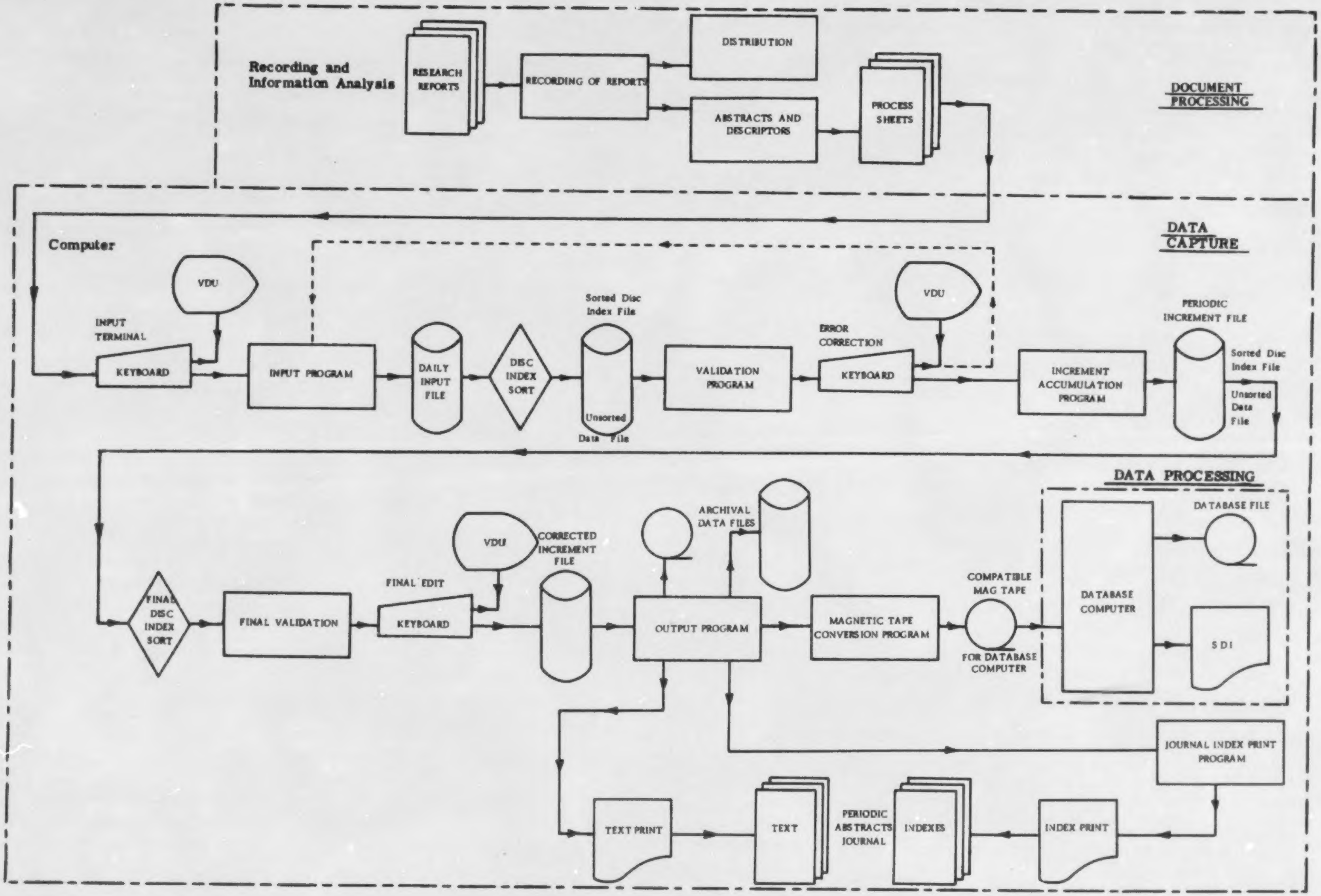

Recording and Information Analysis
RESEARCH REPORTS
RECORDING OF REPORTS
DISTRIBUTION
ABSTRACTS AND DESCRIPTORS
PROCESS SHEETS
DOCUMENT PROCESSING
DATA CAPTURE
Computer
VDU
INPUT TERMINAL
KEYBOARD
INPUT PROGRAM
DAILY INPUT FILE
DISC INDEX SORT
Sorted Disc Index File
Unsorted Data File
VALIDATION PROGRAM
ERROR CORRECTION
KEYBOARD
VDU
INCREMENT ACCUMULATION PROGRAM
PERIODIC INCREMENT FILE
Sorted Disc Index File
Unsorted Data File
DATA PROCESSING
FINAL DISC INDEX SORT
FINAL VALIDATION
FINAL EDIT
KEYBOARD
VDU
CORRECTED INCREMENT FILE
OUTPUT PROGRAM
ARCHIVAL DATA FILES
MAGNETIC TAPE CONVERSION PROGRAM
COMPATIBLE MAG TAPE
FOR DATABASE COMPUTER
DATABASE COMPUTER
DATABASE FILE
SDI
JOURNAL INDEX PRINT PROGRAM
INDEX PRINT
INDEXES
PERIODIC ABSTRACTS JOURNAL
TEXT
TEXT PRINT

Fig.14 Summary of the data preparation and abstract journal production systems at TRC

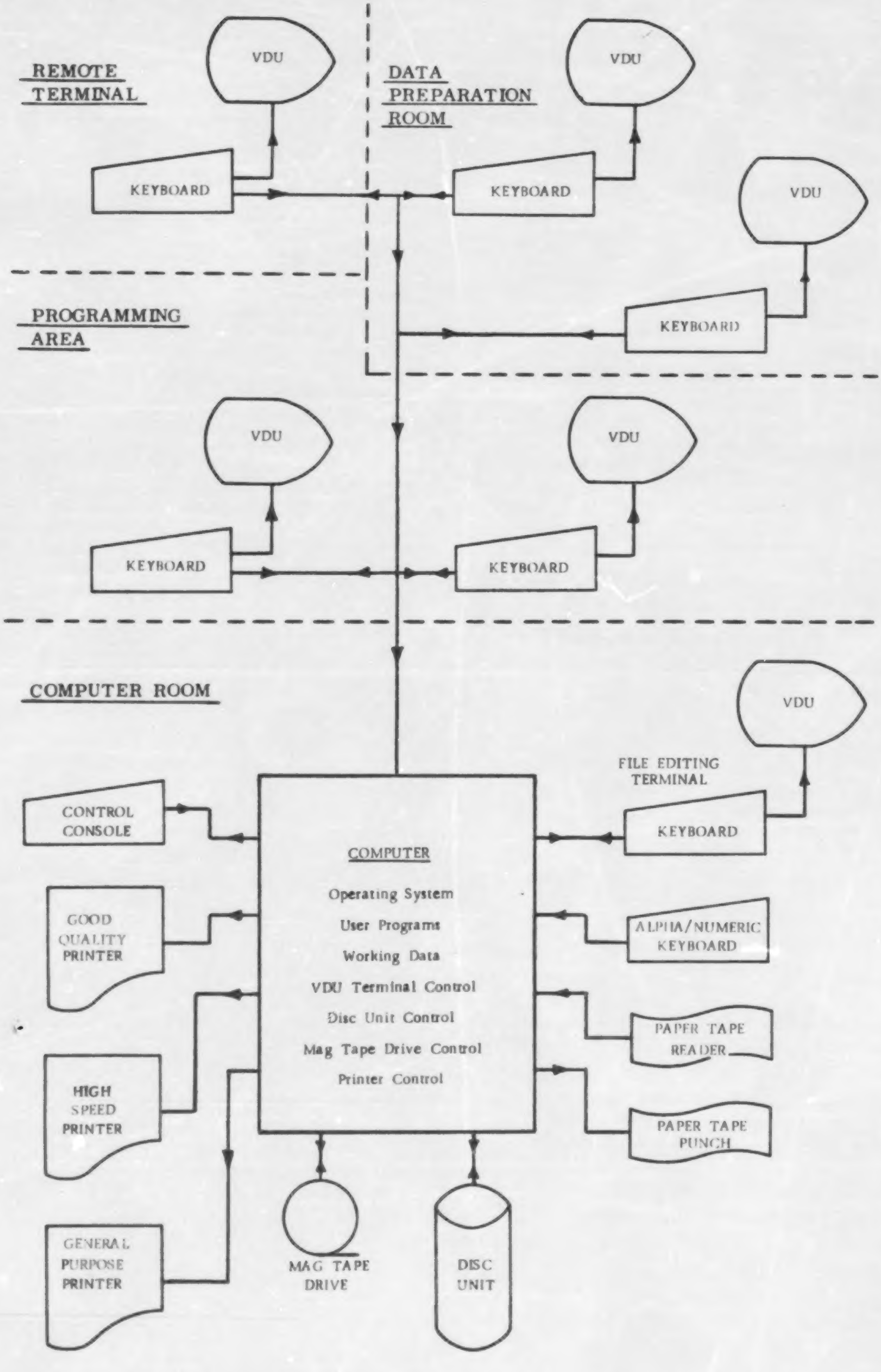

REMOTE TERMINAL
VDU
KEYBOARD
DATA PREPARATION ROOM
VDU
KEYBOARD
VDU
KEYBOARD
PROGRAMMING AREA
VDU
KEYBOARD
VDU
KEYBOARD
COMPUTER ROOM
VDU
FILE EDITING TERMINAL
KEYBOARD
CONTROL CONSOLE
COMPUTER
Operating System
User Programs
Working Data
VDU Terminal Control
Disc Unit Control
Mag Tape Drive Control
Printer Control
GOOD QUALITY PRINTER
HIGH SPEED PRINTER
GENERAL PURPOSE PRINTER
ALPHA/NUMERIC KEYBOARD
PAPER TAPE READER
PAPER TAPE PUNCH
MAG TAPE DRIVE
DISC UNIT

Fig.15 Summary of in-house computer system

the main Computer Room to enable the VDU terminals to be wired directly into the machine. Where the document flow is sufficient to justify the use of a cluster of VDU's it is a good principle to arrange the terminals in pairs so that two adjacent operators can take it in turn to check each others' work.

The recording of the bibliographic data defining each document which enters the system is controlled by means of a Data Entry Form or Document Processing Sheet (see Figure 16). The preparation of the form can be entirely manual or partly by computer as described in paragraph 4.22 which deals with housekeeping and stock control. The form travels to the Data Prep Room with the parent document and, to minimise errors of transcription, as much bibliographic information as possible is copied by the operator direct from the title page of the report. Each day's input is treated as a unit for control purposes, and the Daily Input File is built up on disc as the complete bibliographic data defining each document are keyed in.

4.7 The Data Fields

Referring to Figure 5, it can be seen that most of the data fields are self-explanatory. The general considerations which govern the structure of the data within the Input File are referred to in paragraph 4.2, but some of the fields have special characteristics which are worthy of note.

4.7.1 Field 1 – Main Subject Code. This is the three-character code assigned by the document analyst at the abstracting stage. The code is cross-referenced by the software to a parallel file which contains the master list of the subject headings and sub-headings which has been adopted by the STI system as a whole. The list of COSATI/NTIS codes has been found to be dependable in this context, and the brief but precise statement of subject identity which constitutes Field 1 will be found to have a variety of uses in the STI system (see notes on Field 20).

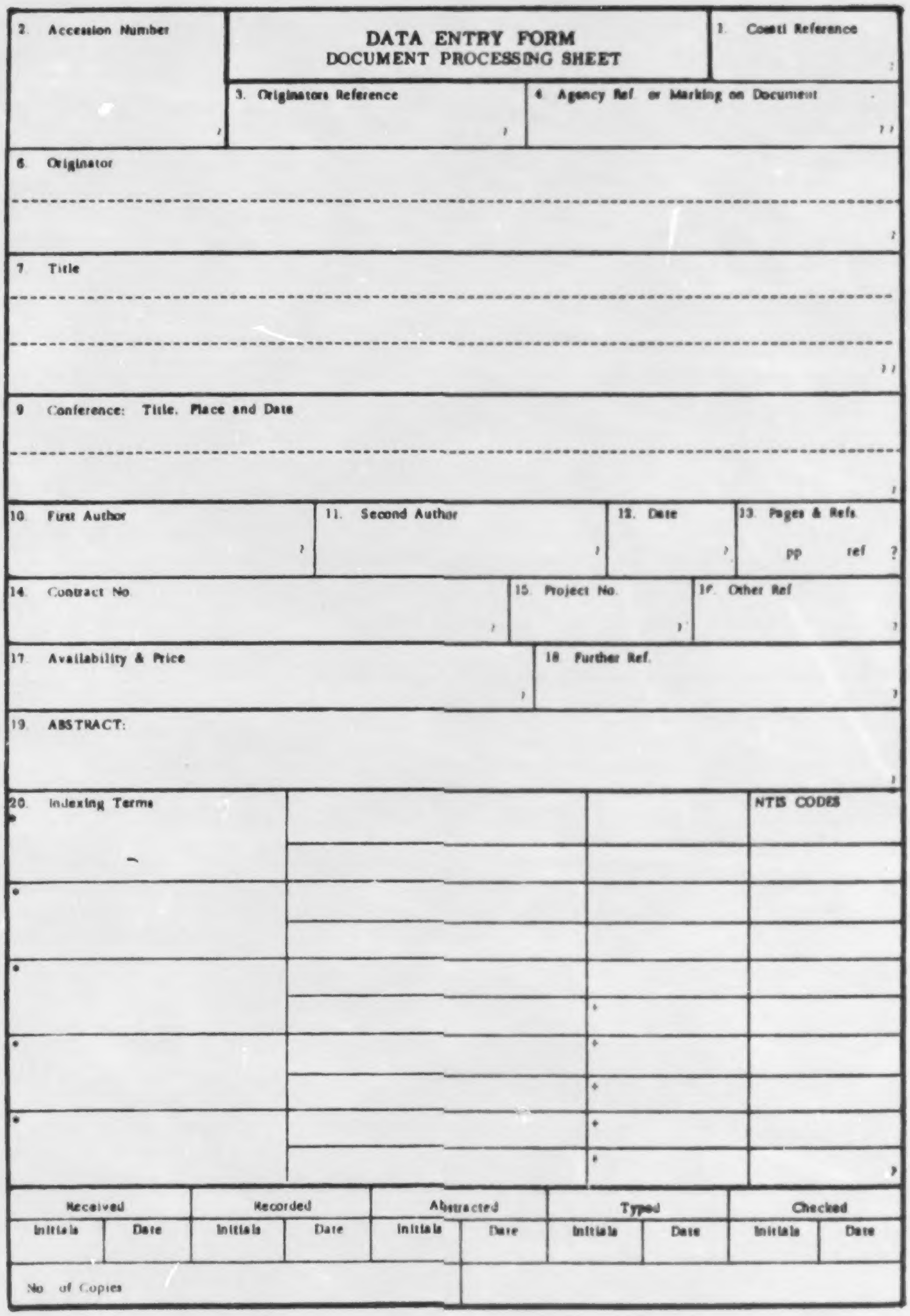
4.7.2 Field 2 – Accession Number. This is the number given to the document when it becomes part of the total store of research report literature. It is usual to employ a system which numbers the documents sequentially within the year of acquisition, so that each annual series begins with the year's identity followed by document Number One.

4.7.3 Field 5 – Increment Number. It can be assumed that the original Feasibility Study (see paragraph 4.1.1) or existing practice will decide the size and frequency of the issues of the STI system's announcement journal. The contents of each issue of the journal will form an increment of the total throughput, and the decision concerning the number of issues to be published in a year will depend upon the quantity of reports handled, the rate at which they are processed, the optimum journal size, etc. Twelve monthly issues may conveniently constitute an annual volume and be numbered accordingly, but it is advisable to introduce multiple volumes in a year if more than twelve issues are to be published in that time. A convenient arrangement is to assign twenty-four or twenty-six issues in a year to two volumes each of twelve or thirteen increments, and the increment number shown at the top right in the example (Fig.10) would be used in such a system (3809? is Increment 9 of Volume 38).

4.7.4 Fields 10 and 11 – First and Second Author. These fields are formatted to suit the requirements of the software which creates the Author Index, i.e. the surname first for ease of sorting into alphabetical order. There need be no restriction on the number of initials allowed provided that the total permitted field size, including spaces and punctuation, is not exceeded.

4.7.5 Field 12 – Date. There are many ways in which dates are traditionally expressed, and the method shown in Figure 10 is a useful compromise. Sufficient dating precision is achieved by quoting only month and year, and the all-numeric presentation is clear and relatively easy for checking by means of the validation software.

4.7.6 Field 20 – Indexing Terms. These can be in the form of keywords (alpha) or subject codes (generally alpha/numeric) and the software should be designed to take account of multi-word terms, provided that they do not exceed the maximum length allowed for the individual descriptors. A maximum length of forty characters has been found to be generally acceptable, only presenting problems to the document analyst when he is dealing with complex chemical compounds. Field 20 is divided into sub-fields (groups of descriptors) by a main sub-field separator which is a double oblique (//) in the example shown in Figure 10, and into sub-sub-fields (individual descriptors) by the secondary separator which is shown as a single oblique (/). This arrangement permits the software to separate the keywords and the subject codes for listing purposes and, at the same time, to give equal status to every term used, be it alpha, alpha/numeric or numeric, for information retrieval purposes. The printed Subject Index will be based upon the terms used in Field 20, and it has been found convenient to designate some or all of the first five descriptors as Subject Headings. The Indexing Terms field begins therefore with up to five descriptors which are asterisked to enable the indexing software to recognise them and copy them over to a special file for use as headings in the Subject Index. Thus they are used both for information retrieval and for printing purposes. There are other special kinds of descriptor such as the "Identifier" which has value for retrieval purposes in relation to a specific report, but is not required in the thesaurus. In the example shown in Figure 9, identifiers are marked with a cross which enables them to be given special treatment by the processing software. Where the COSATI/NTIS system of subject classification is adopted, the first alpha-numeric code following the sub-field separator is designated as the Main Subject Code and is copied to the top of the record (Field 1 in the example shown) where it exercises important control functions in the system software (see paragraph 4.7.1).



DATA ENTRY FORM
DOCUMENT PROCESSING SHEET

2. Accession Number

1. Cosati Reference

3. Originators Reference

4. Agency Ref. or Marking on Document

6. Originator

7. Title

9. Conference: Title, Place and Date

10. First Author

11. Second Author

12. Date

13. Pages & Refs. pp ref

14. Contract No.

15. Project No.

16. Other Ref

17. Availability & Price

18. Further Ref.

19. ABSTRACT:

20. Indexing Terms

NTIS CODES

Received		Recorded		Abstracted		Typed		Checked	
Initials	Date	Initials	Date	Initials	Date	Initials	Date	Initials	Date

No. of Copies

Fig.16 Document processing sheet TRC 1978

4.8 Separator Codes

It is essential that the codes selected to operate as sub-field, field, record and file terminators should never appear as characters in the text of the bibliographic record. A typical function of the sub-field separator (/) has been described in paragraph 4.7.6, and the system shown in Figure 10 uses the question mark (?) as field separator. Records and files are terminated by a single left-hand bracket ([) and a double left-hand bracket ([[) respectively, and care should be taken when choosing these very significant codes to ensure that they can be printed by the system for validation purposes and that they do not have a unique function in the system software provided by the computer manufacturer. A well-designed validation program will test for the presence or absence of separator codes, and will reject any records in which these codes are improperly used. In the system chosen for the purpose of illustration, a record having data in all twenty data fields is a very rare occurrence but, nevertheless, every record must contain twenty field separators in the correct sequence if it is to be accepted. The presence of adjacent question-marks indicates that, for the record shown in Figure 10 there is no information present in those particular fields.

4.9 Keying-in, Validating and Correcting Data

Textually accurate input is essential for the creation of consistent output which will, in turn, provide the user with a good impression of the dependability of the STI system as a whole. It can be deduced from Figure 14 that the data can be validated seven times under the TRC system, twice at the VDU as the data are being captured, twice by software using the Validation Program, and three times from text in the form of computer print-outs which can be created each time a sorted input file has been set up. It is essential that the text editing software should allow items of bibliographic information to be called back to the VDU screen from the Input File in response to keying-in the appropriate accession number. An acceptable working system could be based upon a 2000-character VDU screen, an Input File made up of complete bibliographic records of up to 4000 characters each and a screen format created by the keyboard operator to resemble the layout of records on the ultimate printed page. The keyboard operator, working from the Document Processing Sheet (Fig.16) and the parent document, enters the data in the correct sequence, observing the rules governing the format of each field, and inserts control codes as required. At the end of the record, or having filled the screen with data in cases where the record exceeds 2000 characters in length, the operator reads the contents of the screen and corrects obvious textual errors by use of the edit features of the data prep software. Having completed this first validation, the operator clears the screen by depressing the "SEND" key which consigns the data to the Input File. It is important to ensure that, for records in excess of 2000 characters, the last line of text is retained in the VDU buffer and displayed at the top of the otherwise blank screen when the operator prepares to complete the item by keying-in a second "page" or screen-full of data. This line of text is frequently part of the abstract, and this retention technique enables the operator to maintain continuity without risk of error. The Input Program will be designed to make the data from the two "pages" completely contiguous in the file. Systems employing VDU's with, say, a 1280 character capacity (16 lines each of 80 characters), might have to accept more than two screen-fulls of data per record, depending upon the maximum record size allowed for at the system design stage.

Having incorporated one validation operation into the initial data capture process, the second validation is carried out at the terminal by pairs of operators working as partners. It has already been recommended that input terminals be arranged in pairs and, if two operators spend the first half of the working day independently keying fresh items into the Input File, they can exchange the morning's source documents and spend the second half of the day checking each other's work. This is done by calling back the records from the Input File by their accession numbers and proof reading from the screen. The day's work is then submitted to the Validation Program in the computer which, in a text-handling system, can only carry out elementary checks of the kind indicated in paragraph 4.10. Records which are not acceptable to the program are rejected completely and the related documents are set aside to be keyed in afresh with the next day's work.

To provide the means of carrying out the next two stages of data validation it is necessary to be able to print out the day's work on two-part continuous stationery, preferably using a fast printer which works in upper and lower case, printing all control codes. The data prep supervisor then reads over one copy, looking mainly for consistency of style and record content, while the editor of the announcement journal checks the other, ensuring the accuracy of layout and technical content and also examining the consistency of indexing practice, etc., on the part of the information scientists who analyse the original documents.

Figure 14 refers to a system which creates twenty-four issues of an announcement journal in two volumes, each of twelve issues, in a year. This system, therefore, is based upon a semi-monthly production cycle, and the output of the various related groups within the parent organisation is gauged to fit within this requirement. The implications of this aspect of STI processing are dealt with in detail in paragraph 4.19 but they are referred to at this point in order to emphasize the need for co-ordinating the various activities which make up the overall work pattern. At the end of each data prep cycle a time is allowed for keying-in the final corrections and the incremental input file is then run against the validation program which will list and reject any items which it cannot accept. The next stage is to print out the complete increment on the high-speed printer and submit it to the journal editor who will see that the corrections which he has previously requested have, in fact, been made. The instructions can then be given to print the pages of the journal or to create the magnetic tape which will drive the photo-typesetter.

Seven-stage validation may seem excessive, but the nature of the task is such that opportunity must be given to check that those errors which have been discovered have actually been corrected. It is only necessary for the keyboard operators and the data prep supervisor to proof-read every word of the text in a system of this kind.

4.10 Operator Aids

4.10.1 'Free Typing' Mode – The data prep system so far described only enables the computer to assist the operator at the validation stage, relying almost entirely upon human skill and experience to create records which are correct in every detail. It is a characteristic of text-handling systems that the aids which are provided by the computer can only be minimal and, in the TRC system for instance, they are applied to the main subject code, accession number, increment number, author, date and indexing terms fields. All of these except the author fields are checked for length (individual descriptors will have a maximum length, and the other fields listed above will have a fixed length) and all except the indexing terms field have a fixed format which can be checked by program. Each complete record is scanned to ensure that the correct number of field separators is present, that there are no superfluous spaces in the text, and that the control codes are only placed where they are required to be. The software incorporates a repertoire of error messages which give the operator brief details of any errors found by the computer, and the program rejects faulty records.

4.10.2 'Controlled Typing' Mode – The operator aids described in paragraph 4.10.1 apply to a system which is based upon traditional typing procedures. Such systems offer two advantages in that, by relying on the skill and accuracy of the keyboard operator, the input control software can be greatly simplified and the operator is given a better degree of job satisfaction, generally taking pride in getting things right. It is quite feasible, however, to set up a system which eliminates all formatting requirements at the input stage and expects the operator to simply key in strings of data in response to prompts on the VDU screen. A controlled typing system of this kind might display the names of the various data fields in "inverse video", i.e. dark characters on a white background set up in a local area of the screen, and provide advice concerning the inclusion of control codes (separators, etc) in the text as it is keyed in. The validation software, which must be running continuously, can be designed to apply more extensive checks than are feasible for the free typing mode, and the detection of an error might be signalled by causing the label of the field concerned to flash on the screen. Systems based upon controlled typing of this kind, which fill the input file with unformatted strings of data, require very efficient formatting software to facilitate proof reading and to ensure that the bibliographic records can be displayed to suit the required page layout in the announcement journal.

4.11 Indexes

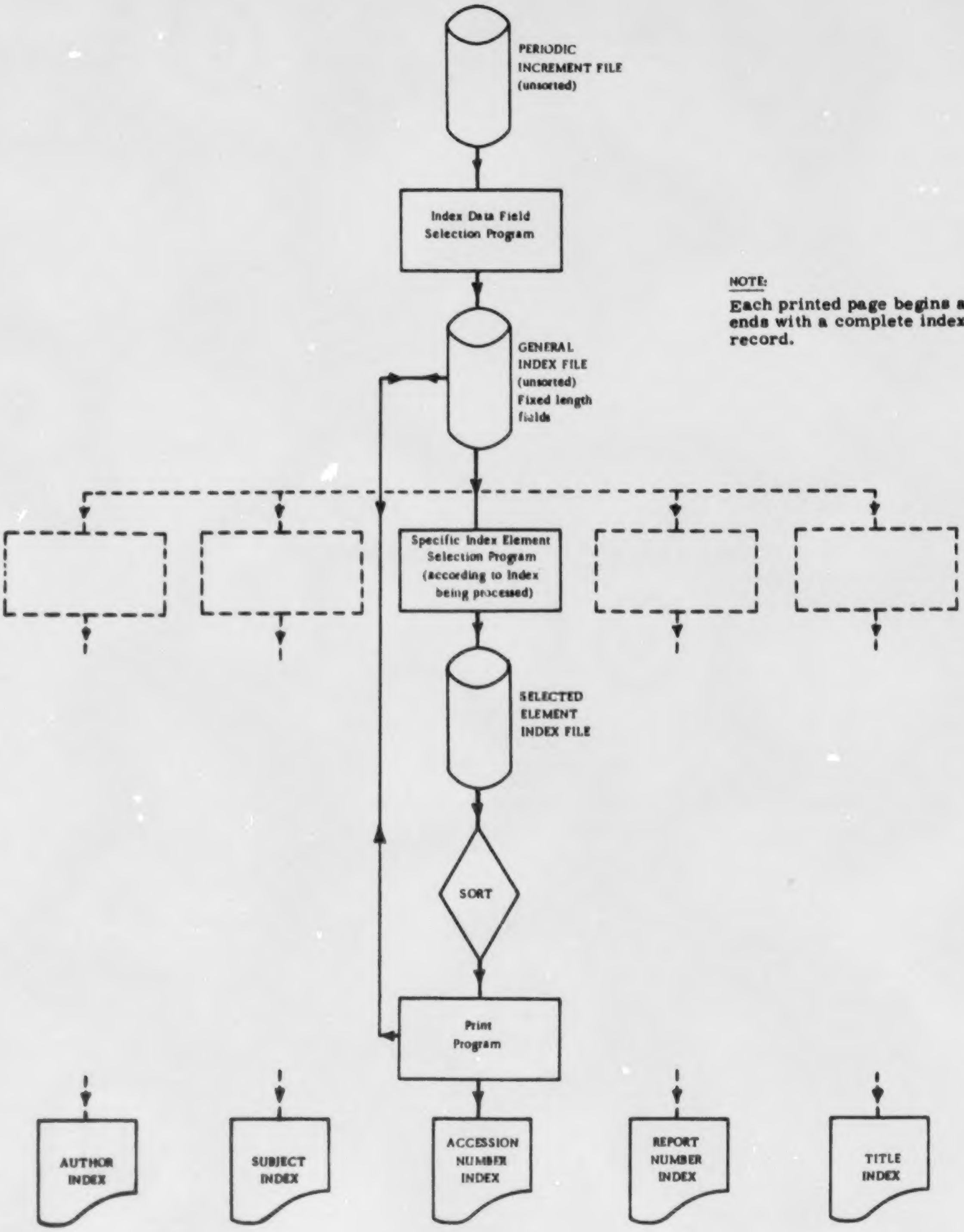
Three main forms of index are generally required for an STI system which prints a journal and operates a data base. These consist of internal machine indexes to facilitate processing, printed indexes for publications and subject indexes stored in the computer for information retrieval purposes.

4.11.1 Machine indexes – The internal indexes used as aids in processing the data are normally invisible to the operator, functioning only within the system software. As records are added to the main disc file a disc index is built up automatically, relating each accession number to the disc address at which its associated bibliographic information is to be found. The data files require frequent sorting and, by use of these indexes, it is only the accession number/address file which needs to be sorted, thus effecting economies in time and resources. When it is necessary for the complete records to be printed or displayed, the software refers to the sorted disc index file which provides pointers to the required bibliographic data in the unsorted main bibliographic file (see Figure 14).

4.11.2 Printed indexes – Figure 17 shows a system of parallel files used for the creation of printed indexes for publications. A typical issue of an announcement journal would be indexed by subject, author, report number and accession number, with an incremental subject code index for control purposes. When volume indexes are produced, combining the incremental indexes over, say, six months or a year, the repertoire might be extended to include title, corporate author and conference indexes.

It is important to observe that, when the input to the system is being treated as a series of data fields in software terms, it is an alpha/numeric sort based on the Main Subject Code field at the beginning of each record which enables the records to be placed in their subject groups. It is this operation which sets up that part of the system which will display the bibliographic file for each increment in printed page sequence. The subject codes are listed in the contents pages at the beginnning of each issue, along with the numbers of the pages upon which they can be found. To locate a complete report reference of which he knows the accession number, the user needs only to be additionally aware of its main subject heading and the number of the increment of the abstract journal in which the item was announced, and he is quickly pointed to the page or group of pages in which it must appear (see Figure 18). The abstracts are arranged on the printed pages in accession number order within subject group.

The report title, its main subject code and the related increment number are the three data fields commonly required in every set of printed indexes, although it may be considered that the title is redundant in the Report Number Index which simply lists all the numbers which refer to a specific report against the main accession number of that report. Again, sorting takes place only on a smaller file when setting up the various indexes (see Figure 17). Each index file is


PERIODIC
INCREMENT FILE
(unsorted)
Index Data Field
Selection Program
GENERAL
INDEX FILE
(unsorted)
Fixed length
fields
Specific Index Element
Selection Program
(according to Index
being processed)
SELECTED
ELEMENT
INDEX FILE
SORT
Print
Program
AUTHOR
INDEX
SUBJECT
INDEX
ACCESSION
NUMBER
INDEX
REPORT
NUMBER
INDEX
TITLE
INDEX

NOTE:
Each printed page begins and ends with a complete index record.

Fig.17 A typical journal index system

CONTENTS

	Page
1. AERONAUTICS	
01A Aerodynamics	1
01C Aircraft	1
01D Aircraft Flight Control and Instrumentation	5
2. AGRICULTURE	
02A Agricultural Chemistry	6
02C Agricultural Engineering	7
02D Agronomy and Horticulture	7
4. ATMOSPHERIC SCIENCES	
04A Atmospheric Physics	7
5. BEHAVIORAL AND SOCIAL SCIENCES	
05A Administration and Management	8
05C Economics	9
05E Human Factors Engineering	11
05H Man-machine Relations	11
05I Personnel Selection, Training, and Evaluation	12
05J Psychology (Individual and Group Behavior)	13
05K Sociology	14
6. BIOLOGICAL AND MEDICAL SCIENCES	
06A Biochemistry	15
06B Bioengineering	15
06C Biology	16
06D Bionics	17
06E Clinical Medicine	18
06H Food	18
06M Microbiology	18

Fig.18 Sample of a contents page (R & D abstracts based on COSATI subject codes)

created automatically from the main bibliographic file, and contains only the principal index elements (the authors' names for the author index, the report titles for the title index, etc) and the disc addresses at which the remainder of the data fields required to make up each index are to be found. When sorting is complete the General Index File, containing all the index information in unsorted form, is run in association with the sorted file of selected index elements to produce the required print-out. The suite of computer programs which creates the printed indexes for each issue of the announcement journal can be so arranged that the programs for the individual indexes are linked together to provide a continuous printing operation and, at the end of the run, the incremental General Index File is retained for archival purposes. A number of these incremental files are combined when it is necessary to create cumulated indexes, but the same software can be used, allowing time for the larger sorting and printing operations, so that the appearance of the index pages remains unchanged.

4.11.3 Indexing for Information Retrieval – Software Features. The systems of indexing used for information retrieval fall into two categories – those used for tracing reports within the stock control files and those used for subject searches. The latter make use of keywords and subject codes for locating report references within the files of bibliographic information which constitute the data base. The stock control files are described in paragraph 4.22 and procedures for information retrieval by subject are generally based upon inverted files of the kind described in Section 7, Vol.III of the Manual. The inverted files used in this context can be briefly defined as bibliographic reference files in which the keyword, and not the record number, identifies the record. Thus a basic inverted file for subject retrieval would consist of a list of the keywords used in the system, with each keyword followed by a set of reference numbers defining the individual records to which that keyword has been allocated. It is essential that the data prep rules concerning the structure of the keywords and the subject codes used for retrieval purposes be carefully observed, and that the separator symbols and other control codes in the Indexing Terms field (Fig.10) be correctly inserted in the records.

The retrieval of report references using the Indexing Term field is relatively straightforward. The inverted index files belonging to the data base are updated as each increment of the announcement journal is added to the store (see Figure 19) so that, against each descriptor in the file, there is a pointer to the location of every posting, i.e. each report reference in the system to which that particular descriptor has been allocated. The subject interest expressed by the user is defined in the form of a list of descriptors linked by Boolean logic ('and', 'or', 'not'), and this search statement is run against the

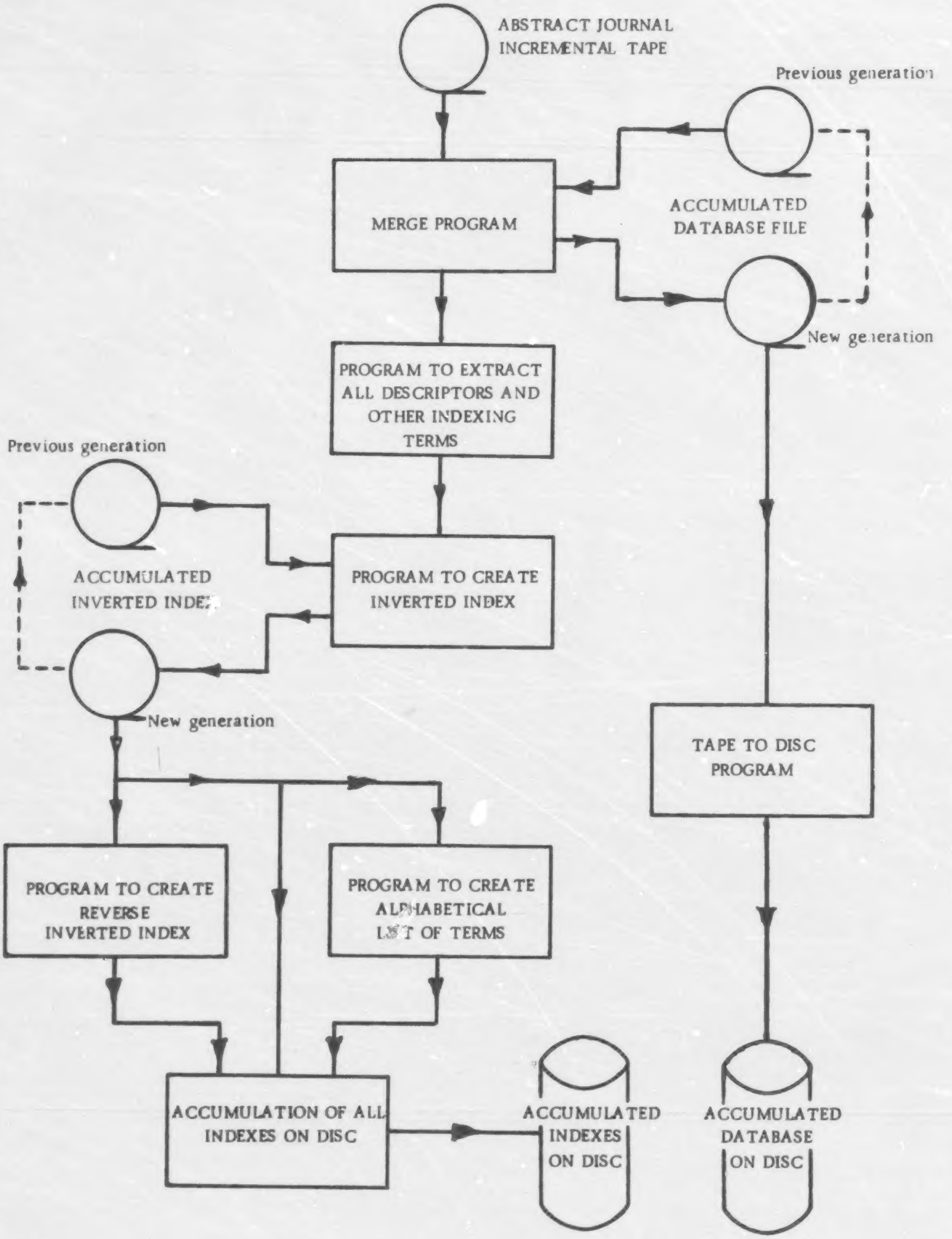

ABSTRACT JOURNAL INCREMENTAL TAPE
Previous generation
MERGE PROGRAM
ACCUMULATED DATABASE FILE
New generation
PROGRAM TO EXTRACT ALL DESCRIPTORS AND OTHER INDEXING TERMS
Previous generation
ACCUMULATED INVERTED INDEX
PROGRAM TO CREATE INVERTED INDEX
New generation
TAPE TO DISC PROGRAM
PROGRAM TO CREATE REVERSE INVERTED INDEX
PROGRAM TO CREATE ALPHABETICAL LIST OF TERMS
ACCUMULATION OF ALL INDEXES ON DISC
ACCUMULATED INDEXES ON DISC
ACCUMULATED DATABASE ON DISC

Fig.19 Creation and update of database files

inverted file. When a match is obtained between a search term and a term in the file, the software codes which act as pointers to the report references linked to the file term are stored as a set within the user file created by the computer for the search, and the total number of these 'hits' is displayed. The sets are manipulated by the computer according to the 'and', 'or', 'not' logic included in the search statement and, as the search proceeds, the user can judge the success or otherwise of his search by the number and quality of the hits held in his latest set. A well-designed system, with large on-line storage facilities and a good response time between the user's terminal and the information store, will allow a searcher to interrupt and redirect his search at any stage, and will display selected bibliographic references as required. It is valuable also to save time and resources by providing a choice of display formats which can be selected by the use of a parameter code keyed in by the user when he gives the print or display command at the conclusion of the retrieval phase of the search. The three optional formats generally offered are (a) accession numbers only, (b) bibliographic information without abstracts but including the descriptor field, and (c) complete records inclusive of abstracts.

Once the inverted file system has been set up, and provided that there is sufficient on-line storage capacity available, it becomes feasible to extend the information retrieval capability to cover data fields in the record other than the descriptor field. Authors, for instance, can be added to the Indexing Term field by software and subjected to straightforward matching logic for retreival purposes. Also, depending upon the resources available, features can be built into the program which will help the searcher to deal with mis-spelled keywords and elementary keying errors.

The Title field can also be automatically extracted from the main bibliographic record, stripped of words of little contextual significance, and set up in an inverted file for information retrieval purposes. Existing software techniques can then be applied to allow the user who wishes to base his search upon two significant terms to either retrieve all report references which contain the two selected terms in their titles or, in more sophisticated systems, to retrieve only those references whose titles contain the two chosen terms in proximity to one another, say up to seven words apart.

Database systems which provide the users with abstracts can, where resources permit, have their retrieval systems extended to apply this proximity logic to all searchable fields including the abstract field. As the technology develops and data storage becomes cheaper it becomes increasingly feasible to hold the text of complete reports in the computer. Proximity logic is a valuable tool for information retrieval in full-text systems.

The information retrieval files so far described deal with complete descriptors, which can be multi-word (with or without hyphens) and/or alpha-numeric, provided that they meet the requirement that the total number of characters, including spaces, within a single descriptor must not exceed the maximum which the system is designed to handle. The machine-readable thesaurus file should be accessible by the user throughout his search and, if he wishes to refer to the thesaurus at any time, he must be able to resume his search at the point at which it was interrupted. The thesaurus file is an example of a structured file, i.e. each descriptor appears as a main term (see Figure 6) followed by its various related terms which are held in their appropriate classes (broader terms, narrower terms, etc). A sub-set of the thesaurus file is maintained which simply lists the main terms in alphabetical order, and this file also should be available to the user when he is carrying out his search.

As well as handling complete descriptors, the system should permit the user to employ truncated terms. A control code which can be printed but which does not appear in normal text can be selected to control this function and, assuming that the asterisk is used for the purpose, appropriate software will interpret 'COMPUT*' as an instruction calling for all terms in the alphabetical list which begin with the six characters 'COMPUT' to be displayed. This facility is known as "back-end truncation" and it provides the searcher with a very useful retrieval tool. "Front-end truncation" is not so straightforward, but it can be achieved by setting up a special version of the alphbetically-arranged file of main terms in which each descriptor is reversed. The user's search terms are similarly reversed before the character comparison logic is applied, so that '*NETWORKS' keyed in at the search terminal can be interpreted by the software as an instruction to display a list of all the descriptors in the thesaurus which end with the characters 'NETWORKS'.

The fact that the total number of postings is stored with each descriptor in the inverted file enables the user to display the number of 'hits' resulting from a search at any stage of his retrieval activity. The user will terminate his search when he achieves a reasonable number of 'hits' of the required quality and, having displayed a few at the terminal for reference purposes, he should be able to print out all his search results off-line at the computer. It is helpful if the searcher can react with the retrieval program in an interactive (conversational) mode so that, for instance, he can be helped from the start to state his search parameters correctly. Towards the end he should be offered the option of the various print or display formats described and finally, the choice of carrying out another search or logging off. The total retrieval activity should be protected by a system of user passwords, and the final 'log-off' instruction should initiate a display on the VDU screen of the time spent in the various files and, where applicable, the cost incurred by the user for his search.

4.12 Key-to-disc Input from Free-standing Terminal Stations

One of the principal applications of computers within the business office environment has come to be known as 'word processing', a term which is generally applied to the printing of correspondence and internal reports by computer-based procedures. Manufacturers of word processing equipment have tended to build free-standing 'intelligent' terminals incorporating a VDU keyboard, a slow printing device capable of producing good quality characters in upper and lower case, an expandable memory and an exchangeable disc unit which is generally based upon a 'floppy' disc (diskette). As an alternative to connecting standard VDU's into an in-house computer, it is feasible to create STI input on word processing terminals and to subsequently transfer the data set up on the exchangeable discs to the larger discs of an in-house machine.

4.13 Key-to-tape Input Systems

A third possibility is the preparation of STI input by the use of shared logic systems. These generally consist of relatively simple VDU keyboard terminals connected by cable to a central magnetic tape unit which incorporates the software which provides 'prompts' to the operators and carries out basic validation. Where more than, say, six terminals are attached to a single tape unit, it is usual to incorporate a supervisor's VDU from which checks can be carried out on the quality and quantity of the input generated by the individual operators. Problems can arise however, due to the fact that magnetic tape files are essentially sequential. It is difficult therefore to make the proof-reading sequence different from the original input sequence, and the selection of individual report references from the tape for display purposes becomes a slow process. The output magnetic tapes however, can be handled by standard software and hardware and each tape has many times the capacity of a floppy disc.

4.14 Input from External Sources

There are many advantages to be gained if individual agencies which process STI by computer make direct contact, both locally and on a world-wide basis, with other organisations carrying out similar activities. It is generally found that individual organisations will adopt internal file formats which will meet local requirements and each can be expected to build its own data base. The most costly part of any comprehensive STI system, in terms of both time and money, is the data preparation and, for organisations which have areas of subject interest in common, effective economies can result from the exchange of magnetic tapes containing ready-formatted bibliographic information. Standards have been laid down for the formats of exchange tapes, but these will generally be found to contain optional fields which require precise definition before they can be accepted by other than the originator's software. Thus it is advisable to enter into direct negotiations with any organisation which offers magnetic tapes on an exchange basis to ensure compatibility of software for data interpretation and conversion.

Two typical bibliographic file layouts are compared in Figure 20 where it can be seen that the conversion program will have to change the sequence of the fields in the originator's tape and omit some of the data completely. The software must also be designed to cope with special characters which may or may not be printable by the recipient's computer, and the allocation of control codes within the originator's system must also be carefully considered.

4.15 Direct Printing via the Computer

The complete STI system which is being described will have an important publishing function, which means that the convenience, speed and quality of printing becomes very significant. The most vital aspect of the various printing operations is the creation of the masters which will provide the pages of announcement journals, but there are many less significant printing jobs which can be satisfactorily carried out on faster machines which may give less perfect outputs. So much of the printed material is in the form of text intended for discerning readers in a library-type environment that output in upper and lower case has come to be regarded as normal. This generally means that the printing peripherals will be operating at less than their maximum design speeds, and this factor must be allowed for at the system design stage. Decisions concerning the printers to suit a specific STI system will depend upon local requirements such as the volume of text to be handled, the standards of quality which have been established and the time constraints with have to be applied at the various stages of production. A very large system producing printed output for subject searches, current awareness runs and regular proof-reading prints might justify the exclusive use of an expensive high-speed drum printer, but greater cost efficiency can often be achieved by sharing the printing load between two or three smaller, slower peripherals. Typical printers suitable for STI purposes are shown in Figures 23 to 26 but bearing in mind the rate at which technological developments are taking place, it is advisable to look closely at the current printer market before making a final choice.

4.16 Printers

Most of the printing devices offered by manufacturers provide ten characters per inch (cpi) across the page and print six lines to the inch, although twelve cpi and eight lines per inch can sometimes be provided. The range of characters and the style of type face to be used in any specific application must be established at an early stage of the design of the system, allowing for the fact that the visual representation of control codes, such as field and record separators, will be required from time to time. Sets of 64 different characters are usually offered for text applications in which no lower case alphabet will be printed, and 96 character sets are recommended wherever output is required in both upper and lower case. Two typical 96 character sets are shown in Figures 21 and 22 and it can be seen that symbols are available to represent a wide range of control codes and other characters which are normally "non print". Printing devices suitable for STI applications can be grouped into the four main categories which are described briefly below, although it can be expected that individual manufacturers will offer special features within these broad classes.

DATA BASE FILE		INPUT FROM EXTERNAL SOURCE	
Field Number	Contents	Field Number	Contents
1	Subject reference		
2	Accession number	001	Record number
3	Originator's reference	066	Microfiche code
4	Agency's reference number	105	Volume number
5	Project number		
6	Reference number		
7	Further reference number	106	Issue number
8	Contract number		
9	Date of report	113	Publication date
10	Pagination	112	Pagination
11	Author 1	109	Authors
12	Author 2	109	Authors
13	Availability and price		
14	Originator code		
15	Originator entry	102	Coden
16	Title in English	108	First title
17	Title in original language	108	Second title
18	Conference details	131	Originating conference
19	Abstract		
20	Indexing terms	572	Posted descriptors

Fig.20 Harmonisation of data base files

4.16.1 The Drum Printer (Fig.23) is the fastest kind of printing peripheral in general use, with many manufacturers offering speeds of about 2000 lines per minute. Characters are embossed upon the surface of the rotating drum and fixed hammers, one for each character position along the length of the drum, force the paper instantaneously onto the faces of the selected characters. This sharp, repeated impact, applied via the inked ribbon, creates lines of characters on the printed page, so that the device is also known as a 'line printer'.

4.16.2 The Matrix Printer (Figures 22 and 24) creates characters from a matrix of dots. Manufacturers offer ingenious ways of minimising the spaces between dots to give the impression of a continuous ink line, and it follows that the closer the proximity of the dots (dots per inch) within the matrix, the denser and more attractive will be the characters produced. A minimum of seven dots is recommended across the face of each character, and a minimum of nine dots is required for each character vertically to allow for the production of true descenders, ie the stems of 'g', 'j', 'p', etc. The characters, which are provided in the usual sets, are created by means of a sequence of impulses generated within the machine (hard-wired logic) and these are converted into patterns of dots on the printed page according to any one of a variety of mechanical processes. The increasing availability of micro-processors makes it much more feasible to inject special characters into the printing sequence under the control of the user's print program. Matrix printers are generally slower and less expensive than drum printers, providing good quality output at speeds between 300 and 600 lines per minute.

4.16.3 The 'Daisy Wheel' Printer (Fig.25) produces output of very good quality at speeds around 50 characters per second, approximating to 30 lines per minute. A repertoire of 88 characters is generally provided for upper and lower case applications, with the characters embossed at the ends of 'petals' formed by means of radial cuts in a flat disc about 75 mm in diameter. As the disc rotates the selected characters are pressed onto the paper, through the inked ribbon, to create lines of print. Printers of this kind have very few moving parts and are the least expensive to buy, and they have the advantage of providing a variety of type faces according to the disc selected.

Char.	Octal	Hex
A	0404 101	41
B	0410 102	42
C	0414 103	43
D	0420 104	44
E	0424 105	45
F	0430 106	46
G	0434 107	47
H	0440 110	48
I	0444 111	49
J	0450 112	4A
K	0454 113	4B
L	0460 114	4C
M	0464 115	4D
N	0470 116	4E
O	0474 117	4F
P	0500 120	50
Q	0504 121	51
R	0510 122	52
S	0514 123	53
T	0520 124	54
U	0524 125	55
V	0530 126	56
W	0534 127	57
X	0540 130	58
Y	0544 131	59
Z	0550 132	5A

Char.	Octal	Hex
0	0300 060	30
1	0304 061	31
2	0310 062	32
3	0314 063	33
4	0320 064	34
5	0324 065	35
6	0330 066	36
7	0334 067	37
8	0340 070	38
9	0344 071	39
Space	0200 040	20
!	0204 041	21
"	0210 042	22
£or#	0214 043	23
$	0220 044	24
%	0224 045	25
&	0230 046	26
'	0234 047	27
(	0240 050	28
)	0244 051	29
*	0250 052	2A
+	0254 053	2B
,	0260 054	2C
–	0264 055	2D
.	0270 056	2E
/	0274 057	2F

Char.	Octal	Hex
\	0600 140	60
a	0604 141	61
b	0610 142	62
c	0614 143	63
d	0620 144	64
e	0624 145	65
f	0630 146	66
g	0634 147	67
h	0640 150	68
i	0644 151	69
j	0650 152	6A
k	0654 153	6B
l	0660 154	6C
m	0664 155	6D
n	0670 156	6E
o	0674 157	6F
:	0350 072	3A
;	0354 073	3B
<	0360 074	3C
=	0364 075	3D
>	0370 076	3E
?	0374 077	3F
@	0400 100	40
[	0554 133	5B
½	0560 134	5C
]	0564 135	5D
∧	0570 136	5E
└	0574 137	5F

Char.	Octal	Hex
p	0700 160	70
q	0704 161	71
r	0710 162	72
s	0714 163	73
t	0720 164	74
u	0724 165	75
v	0730 166	76
w	0734 167	77
x	0740 170	78
y	0744 171	79
z	0750 172	7A
{	0754 173	7B
|	0760 174	7C
}	0764 175	7D
¬	0770 176	7E
DEL	0774 177	7F

Fig.21 The 96-character set used by the BCL M18 computer

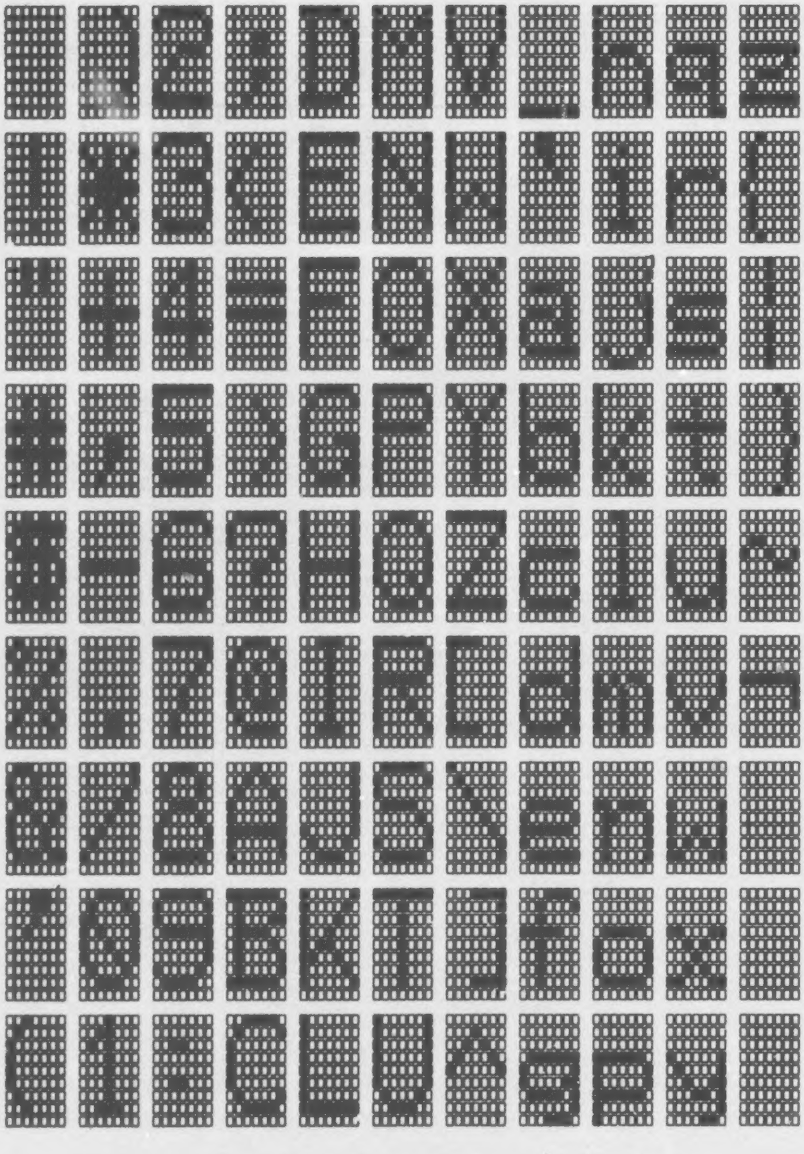


Fig.22 A typical 96-character set.
9 x 7 dot matrix with true descenders

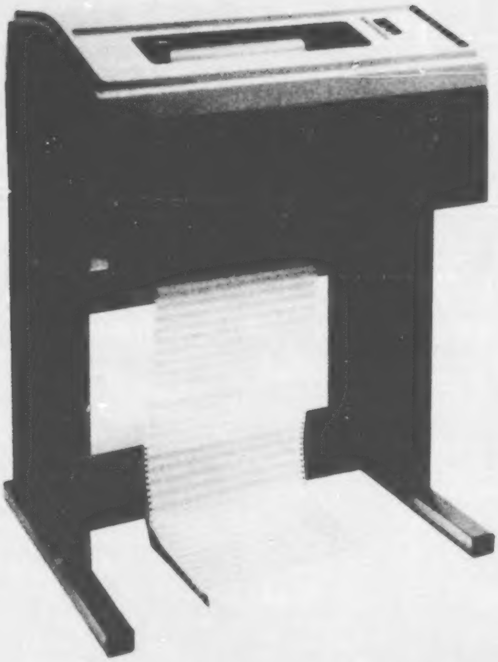


Fig.23 The Dataproducts 2230 drum printer

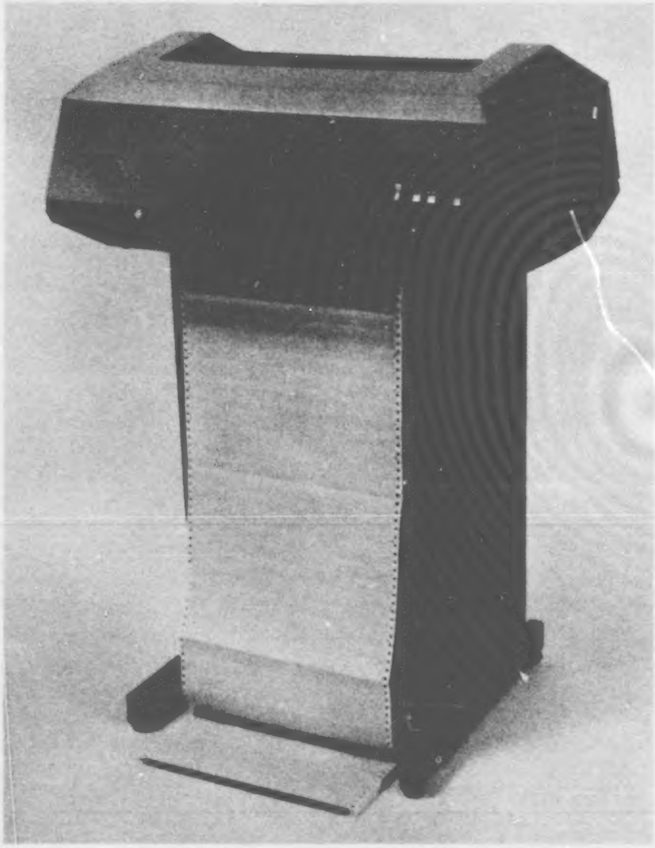


Fig.24 The Tally 3000-series matrix printer

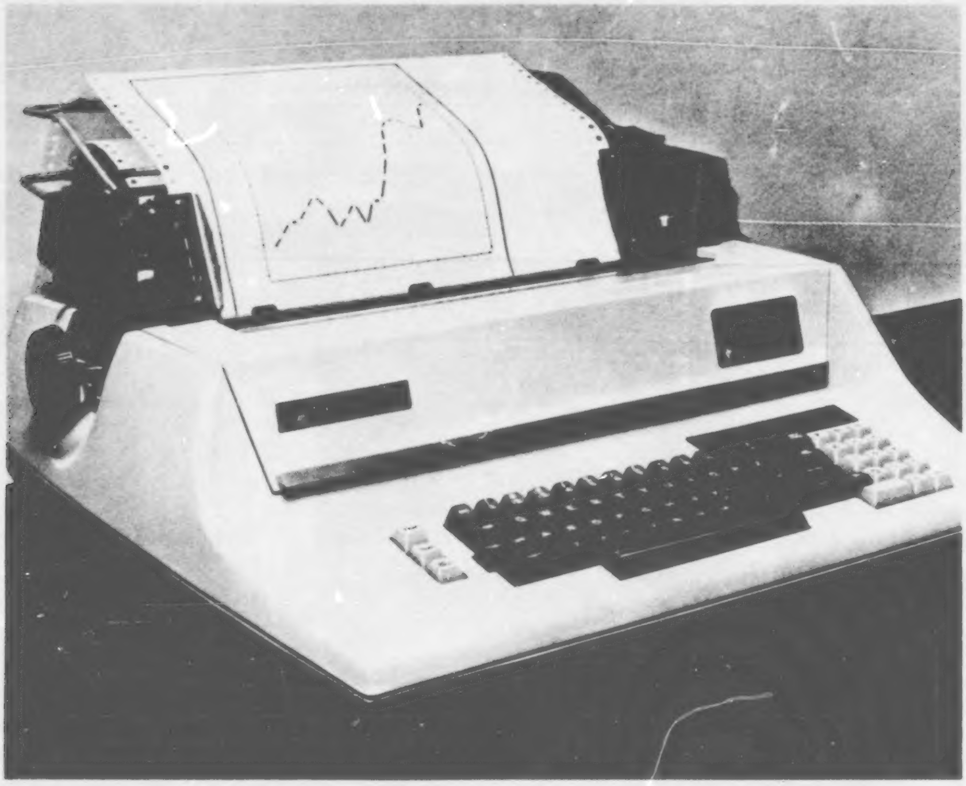


Fig.25 The Diablo 1620 petal printer

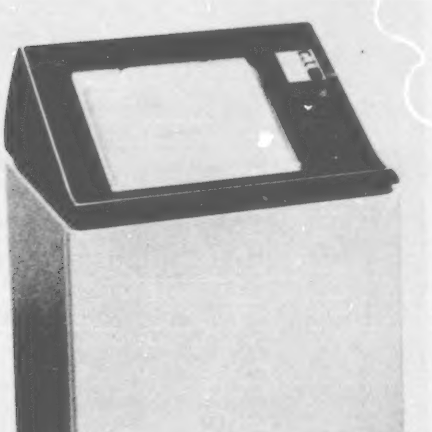


Fig.26 The Versatec electrostatic printer

4.16.4 The Electrostatic Printer (Fig.26) is relatively expensive, but it provides output of the best quality. The three kinds of printer described so far can give multiple copies at a single pass by using continuous stationery inter-leaved with carbon or with paper especially treated to provide multi-part prints as a result of an impact process but without carbons. The electrostatic printer however, provides single copies only on paper which is sensitised to record electrical impulses. Each character is created initially as a pattern of electrically charged dots and the paper is then brought into contact with magnetic ink which adheres only to the charged areas. The ink is dried instantaneously and a permanent record is retained on the paper. Hard-wired logic boards provide a range of character sets, and the process allows for the dots to be more closely packed than is possible with matrices which are created mechanically. Characters appear to be more intensely black and are thus more attractive, and economies in the use of paper can be achieved in some instances by operating at more than twelve characters per inch. Using normal characters electrostatic machines can provide 1000 lines of print per minute, but the process is closely allied to graph plotting so that it can give an almost limitless range of characters and character sizes at slower speeds in response to appropriate logic in the user's program. Some manufacturers provide sets of cyrillic characters which are generated from logic boards plugged into the machine.

4.17 The Remote Printing of Computer Output

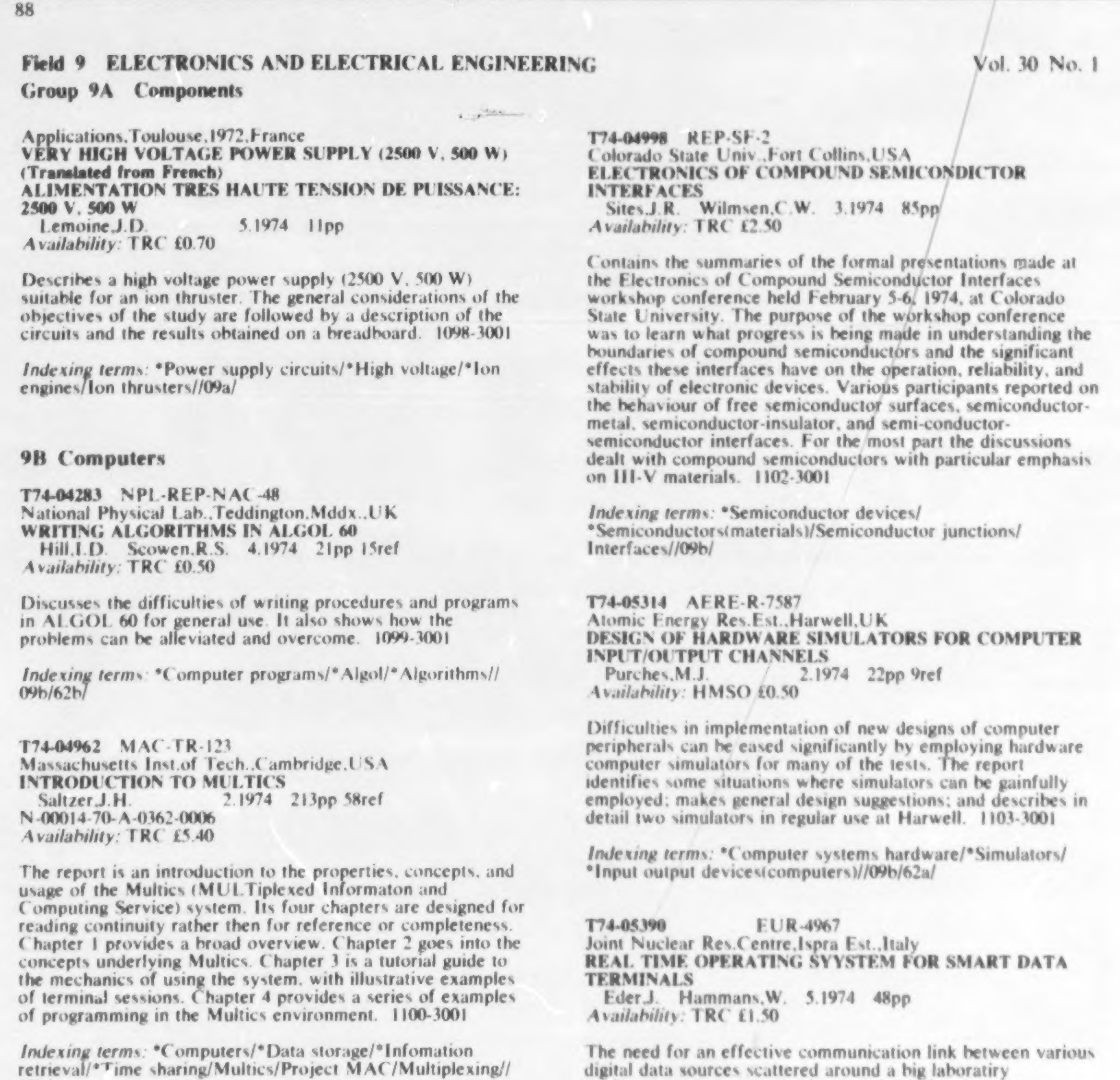
It has been assumed so far that the printing devices for the production of computer output in the form of paper-copies (hardcopy) are linked directly to the processor. The buffers into which the data are fed from the processor may only hold one or two lines of characters at a time, so the speed of operation becomes very slow when compared with the data transfer rates of magnetic media (tapes or discs) or even of punched paper tape in some instances. Where the demand for printed output justifies it, greater overall efficiency can be achieved by the use of remote printing devices, and the introduction of such 'off-line' resources opens the way to the consideration of computer-controlled photo-typesetting and a standard of page presentation which can compare favourably with that of a well-printed book. If it is decided, in the course of the design of an STI system, that printing will take place off-line, then the end product of the data processing stage of the operation will be in the form of punched paper-tape, magnetic tape or floppy disc, each especially formatted to meet the requirements of the selected printing device.

4.17.1 Off-line printing can be used in association with any of the printers already described and is recommended as an efficient way of dealing with long and uninterrupted print jobs. In a time-sharing computer system, i.e. one in which several programs can be run simultaneously, the print program for the production of, say, the pages of the announcement journal can be run as a background job very economically, but a good quality printer must be dedicated to that job throughout the run. A magnetic tape set up for printing in this way is valuable for archival purposes and will also provide the input for any other incremental tape which may be required, e.g. the database update tape or an information exchange tape.

4.17.2 Computer-controlled photo-typesetting provides printout of a very high quality but involves rather costly equipment. As the name implies the basic output of the photo-typesetter is an image, generally a page or column of text, created on photo-sensitive paper by exposing it under strictly controlled conditions to light or an electron beam which has passed through a translucent medium on which characters are inscribed. Some modern devices use electronic character generation techniques with a cathode ray tube display. Type sizes can be varied by means of lenses, and a single character grid or a specific set of character generation logic is generally of a sufficient capacity to accommodate several related alphabets, thus providing a number of different type faces. The device responds to two data streams, the first of which selects or creates the characters, leaving the second to place them in their proper sequence on the display tube for subsequent exposure, line by line, onto photosensitive paper or film. A Linotron 505 photo-typesetter was used to create the page of text shown in Figure 27 operating upon data supplied on magnetic tape. Figure 28 illustrates the operating principles of the Linotron 202 machine, and typesetters of this kind provide film output of a size which is directly commensurate with the normal printed page. Other designs of photo-typesetter produce microfilm output (computer output microfilm or 'COM' – see Section 9, Vol.III of the Manual) which must be enlarged photographically when printed pages are required. Logic can be devised to enable the larger machines to carry out automatic hyphenation of words at the end of lines of text and to create a uniform line length by varying the spaces between words and characters, a feature known as right-hand justification. Simpler and slower type-setters use character sets inscribed upon continuous bands of film, several of which can in some instances be accommodated simultaneously in the machine to give a range of type faces. Depending upon the type faces used and any special format features involved, it can take just a few minutes to create a typset page. Programming for computer-based photo-typesetting requires special skills, since the techniques involved are expressed in terms familiar to the printing industry, and the output requires expert line-by-line analysis in order to lay down the software logic rules which will govern the format to be adopted for the display of the data. The process can employ type faces which are both small and clear, and can space the characters proportionately in the line, thus almost doubling the quantity of text on a page as compared with the output from conventional computer printers.

4.18 The Layout of Printed Pages

In order that the maximum benefit may be obtained from an STI system it is essential that the best quality output shall be that which is read, either on a VDU screen or in a printed publication, by the user. Extensive work has been carried out by the UK Royal College of Art on the readability of print, and the appearance of the printed page, both from the practical and the aesthetic points of view, is of great significance to the user of the system when he assesses the acceptability of its products. The means available for creating well laid-out pages are therefore described in detail in the



88

Field 9 ELECTRONICS AND ELECTRICAL ENGINEERING Vol. 30 No. 1

Group 9A Components

Applications,Toulouse,1972,France
VERY HIGH VOLTAGE POWER SUPPLY (2500 V, 500 W) (Translated from French)
ALIMENTATION TRES HAUTE TENSION DE PUISSANCE: 2500 V, 500 W
Lemoine,J.D. 5.1974 11pp
Availability: TRC £0.70

Describes a high voltage power supply (2500 V, 500 W) suitable for an ion thruster. The general considerations of the objectives of the study are followed by a description of the circuits and the results obtained on a breadboard. 1098-3001

Indexing terms: *Power supply circuits/*High voltage/*Ion engines/Ion thrusters//09a/

9B Computers

T74-04283 NPL-REP-NAC-48
National Physical Lab.,Teddington,Mddx.,UK
WRITING ALGORITHMS IN ALGOL 60
Hill,I.D. Scowen,R.S. 4.1974 21pp 15ref
Availability: TRC £0.50

Discusses the difficulties of writing procedures and programs in ALGOL 60 for general use. It also shows how the problems can be alleviated and overcome. 1099-3001

Indexing terms: *Computer programs/*Algol/*Algorithms// 09b/62b/

T74-04962 MAC-TR-123
Massachusetts Inst.of Tech.,Cambridge,USA
INTRODUCTION TO MULTICS
Saltzer,J.H. 2.1974 213pp 58ref
N-00014-70-A-0362-0006
Availability: TRC £5.40

The report is an introduction to the properties, concepts, and usage of the Multics (MULTiplexed Informaton and Computing Service) system. Its four chapters are designed for reading continuity rather then for reference or completeness. Chapter 1 provides a broad overview. Chapter 2 goes into the concepts underlying Multics. Chapter 3 is a tutorial guide to the mechanics of using the system, with illustrative examples of terminal sessions. Chapter 4 provides a series of examples of programming in the Multics environment. 1100-3001

Indexing terms: *Computers/*Data storage/*Infomation retrieval/*Time sharing/Multics/Project MAC/Multiplexing// 09b/

T74-04963
Arizona Univ.,Engineering Experiment Station,Tucson,USA
INTERACTIVE GRAPHICS SOFTWARE FOR USE IN FINITE ELEMENT ANALYSIS (ANNUAL REPORT)
Kamel,H.A. McCabe,M.W. 2.1974 176pp 30ref
N-00014-67-A-0209-0016
Availability: TRC £4.40

Includes a summary of activities during the 1973 calender year, as well as a number of appendices related to the development of the GIFTS (Graphic oriented Interactive Finite element analysis on Time Sharing) system during this period. Appendix A contains a paper outlining the current GIFTS system. This paper has been submitted for presentation to the ONR International Symposium on Structural Mechanics Software, Maryland, June 1974. One of the test cases chosen for the system was a two-dimensional analysis of a human bone in support of a theory linking bone stresses to arthritis of the joints. This paper is included as Appendix B. A revised and updated user's manual and the newest format for the Unified Data Base are included as appendices. A new item is a solved example manual containing two problems of varying complexity. Additional appendices pertaining to the improvement of solution efficiency are also included. 1101-3001

Indexing terms: *Computer systems programs/*Graphic methods/*Time sharing/Visual aids/Ship structural components/Structural design/Bones/Bioengineering/Computer aided design/GIFTS program//09b/

T74-04998 REP-SF-2
Colorado State Univ.,Fort Collins,USA
ELECTRONICS OF COMPOUND SEMICONDICTOR INTERFACES
Sites,J.R. Wilmsen,C.W. 3.1974 85pp
Availability: TRC £2.50

Contains the summaries of the formal presentations made at the Electronics of Compound Semiconductor Interfaces workshop conference held February 5-6, 1974, at Colorado State University. The purpose of the workshop conference was to learn what progress is being made in understanding the boundaries of compound semiconductors and the significant effects these interfaces have on the operation, reliability, and stability of electronic devices. Various participants reported on the behaviour of free semiconductor surfaces, semiconductor-metal, semiconductor-insulator, and semi-conductor-semiconductor interfaces. For the most part the discussions dealt with compound semiconductors with particular emphasis on III-V materials. 1102-3001

Indexing terms: *Semiconductor devices/ *Semiconductors(materials)/Semiconductor junctions/ Interfaces//09b/

T74-05314 AERE-R-7587
Atomic Energy Res.Est.,Harwell,UK
DESIGN OF HARDWARE SIMULATORS FOR COMPUTER INPUT/OUTPUT CHANNELS
Purches,M.J. 2.1974 22pp 9ref
Availability: HMSO £0.50

Difficulties in implementation of new designs of computer peripherals can be eased significantly by employing hardware computer simulators for many of the tests. The report identifies some situations where simulators can be gainfully employed; makes general design suggestions; and describes in detail two simulators in regular use at Harwell. 1103-3001

Indexing terms: *Computer systems hardware/*Simulators/ *Input output devices(computers)//09b/62a/

T74-05390 EUR-4967
Joint Nuclear Res.Centre,Ispra Est.,Italy
REAL TIME OPERATING SYYSTEM FOR SMART DATA TERMINALS
Eder,J. Hammans,W. 5.1974 48pp
Availability: TRC £1.50

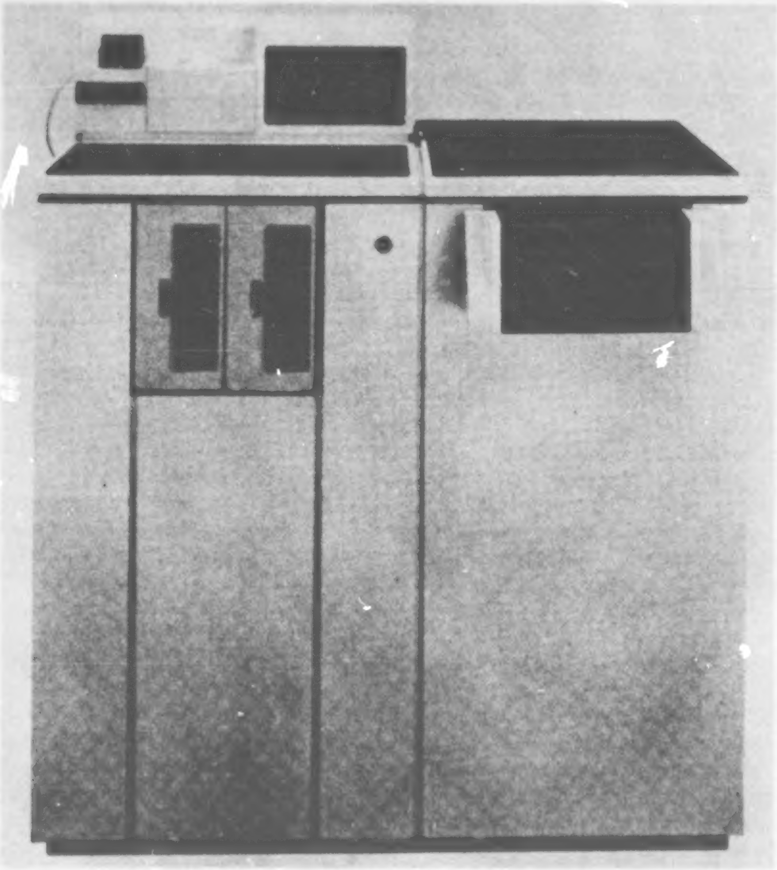
The need for an effective communication link between various digital data sources scattered around a big laboratiry environment and a commonly shared data logging system (SPC-16 minicomputer with disc and magtape) led to the development of SMART datastations (1). SMART datastations are active, they can take over partially control of the computer. The terminals' user has real time access and may influence the data flow in the computer. A software package is described in this report which allows to control several simultaneously working terminals in a disc based operating system (DBOS-16). A control routine recognises a set of control commands from the terminals and stores incoming data due to the respective station address on disc. Some off-line routines are provided for file assignment, code conversion and editing of IBM compatible magnetic tapes. 1104-3001

Indexing terms: *Real time operations/*Telemettry/ *Minicomputers,*Computer systems programs/SMAART/ Data processing//09b/

T74-05397 EUR-5121
Gemeinsame Kernforschungsstelle Forschungsanstalt-Ispra, Italy
INTERFACE FOR THE CONTRO OF A MULTICHOPPER FACILITY (Report in German)
INTERFACE ZUR STEUERUNG EINER MULTICHOPPERANLAGE
Eder,J. Brak,H. 5.1974 22pp 2ref
Availability: TRC £1.00

An interface for the coupling of a multi-chopper facility to a computer has been developed, which provides program controlled settings of rotor speed (period time) and phase

Fig.27 Computer typeset abstract journal
— A sample page.



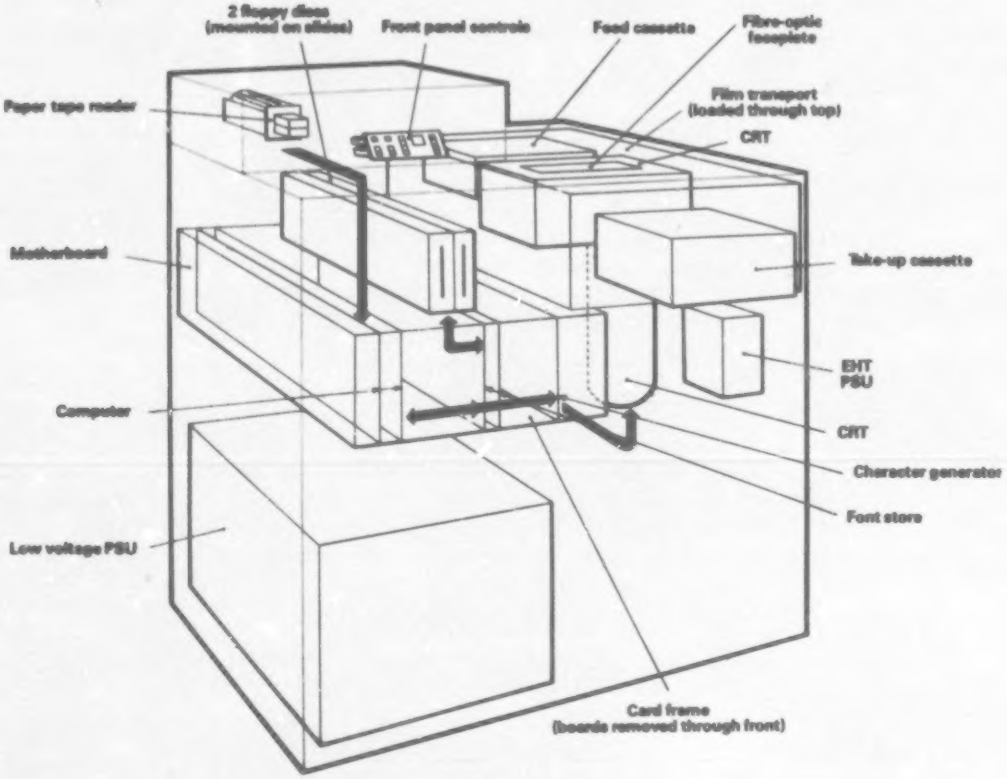

2 floppy discs (mounted on slides)
Front panel controls
Feed cassette
Fibre-optic faceplate
Paper tape reader
Film transport (loaded through top)
CRT
Motherboard
Take-up cassette
EHT PSU
Computer
CRT
Character generator
Font store
Low voltage PSU
Card frame (boards removed through front)

Fig.28 The Linotron 202 computer-controlled photo-typesetter

ensuing paragraphs, for there are features which can be incorporated into the printing procedures to give output formats which are both effective and pleasing to the eye as well as reducing fatigue in the reading of the text. It is usual to have a relatively slow printing device available for high-quality work, but allowance must also be made for those parts of the system which produce programs listings, operating statistics and similar routine printed output for use within the computer area only. Such work demands speed and efficiency rather than quality of print, provided that it is legible.

4.18.1 Announcement Journals and Accession Lists require special care, and there are four important features which combine to create a good page format. If the pages are arranged according to subject and the main subject headings are in alphabetical order throughout the publication, each left-hand page of text should display at the top, printed bold, the main subject heading assigned to the first item of bibliographic information on that page, and each right-hand page should display the heading assigned to the last item on that page. Even if there are items belonging under several headings on a pair of facing pages, it is sufficient to display only the first and the last. If adjacent pages contain only items belonging to the same main subject group, then the page heading should be followed by the word '(continued)', bracketed as shown, wherever it applies. Secondly, the accession number which uniquely identifies each item in the publication should stand out boldly. Thirdly, there should be a sufficient allowance of 'light', i.e. space between words and between lines on each page, without wasting paper and, finally, no item should start too near the end of a column or page of text. The impression of bold type can be created on some printers by holding a line of text in a buffer and repeating the line by using the 'Carriage Return' code without a 'Line Feed'. Otherwise upper case characters used in conjunction with careful spacing can be quite effective. Fourthly, the print program should be designed to look ahead to avoid starting any items too near the bottom of a column or page of text. A useful standard is the 'five-line rule' whereby there must be room to accommodate at least five lines of text before a fresh item is started at the end of a page. The routines set up for printing main subject headings and sub-headings must also be taken into account when programming to incorporate this rule. It would be advantageous to hold the accession number belonging to each item in store throughout the printing of the whole item. Then, should it prove necessary to continue the item on a second page, the heading information on the new page could be followed by a repeat of the accession number followed by the word '(continued)' placed after it in brackets as shown.

It is usual to hold the main subject headings and the sub-headings in both coded form and in text in a look-up list which resides in the computer's data store when the printing program is being run. The first data field in each record should be the main subject code which is matched by program against its counterpart in the look-up list. This matching operation is then used as a trigger to copy the complete heading from the list into its proper position on the printed page, thus basing this part of the system upon an authority list which can be readily updated when additional scientific and technological disciplines have to be accommodated. Page numbers can also be held in a look-up list controlled by program logic which will place the odd numbers in the top right-hand corner of right-handed pages and the even numbers in the top left-hand corner of left-handed pages. Page 1 of a publication should always be a right-hand page.

4.18.2 Photo-reduction. It is quite feasible to create two or three-column page structures (two-up or three-up pages) when printing by computer-controlled photo-typesetter, but it is wise to use only single column pages in systems which employ orthodox printers. Assuming standard A4 sizes with a page width of 8¼ inches (210 mm), the text can occupy up to 7 inches (178 mm) of lie length to give a binding margin of 7/8 inch (22 mm) and a free margin of 3/8 inch (10 mm) along the opposite edge. Where no photo-reduction facilities are available it may be necessary to base the page structure on a 70 character line (7 inches x 10 characters to the inch), but it can be more satisfactory and economical to allow the full 80 character line, which is generally available on the VDU screens used at the data prep stage, to form the basis of the printed output. There is a wide range of equipment available on the commercial market from companies such as 3M's and Kodak which will provide variable photo-reduction ratios, some offering a range of sizes from 50% reduction to 50% enlargement of the original. The facility for reducing original sizes has been proved to be the more useful, and well-produced 10 character to the inch computer printer output can be quite legible down to about 60% of the original character size, i.e. with the camera set to enable a line of 120 characters occupying 12.0 inches (304 mm) to fit into about 7 inches (178 mm). As indicated in paragraph 4.18.3 this degree of photo-reduction can prove particularly valuable in the production of indexes to announcement journals since these are generally browsed through, whereas the main pages of text are more carefully read and therefore require the best possible quality of print.

If it is decided to produce the master pages of text for publications by photo-reducing original output, then the computer printer will be expected to create especially good camera-ready copy. The 12½% reduction required to place 80 characters in a 7 inch (178 mm) line length will be automatically applied to the length of the page as well, but it is generally considerations of width which govern page sizes, since it is relatively easy to adjust the line-counts in a print program. There must be an allowance of space at the top and bottom of each page, so, assuming six lines to the inch on the original printer output and a final page length of 11¾ inches (297 mm), each page can be allowed to contain a maximum of 65 lines. This figure is established by the fact that the continuous stationery for computer printers normally measures 11 inches (278 mm) between transverse perforations and, on the photo-reduced page, 64 lines can give a 1 inch (25 mm) clear margin at the top and 1¼ inch (32 mm) clear at the bottom. The positioning of the page numbers and the allocation of blank lines to give light to a page are important features in the design of good printing formats.

4.18.3 Printed indexes do not generally require the same high standards of printing and page layout as are necessary for the main text. The sample shown in Figure 3 was produced in upper case only using a 120 character position line and allowing a two page length of continuous stationery (22 inches or 556 mm) per page of text, printing a maximum of

108 lines (18 inches or 457 mm). This process requires careful setting up of the paper on the computer printer to ensure that the transverse perforations pass between two adjacent lines of print, and the final page master is obtained by photo-reducing the computer output sheets to 56% of their original size. The sample page shown in Figure 29 has been photo-reduced to 70% of the original size on a fixed ratio offset-litho plate making camera device. It is necessary to make careful calculations when defining the original line and page lengths for examples of this kind before laying down the print program parameters, and it is advisable to show samples of the photo-reduced output to members of the staff of the Centre and to invite their comments on the acceptability of the print. The legibility of the text shown in Figure 29 might be improved, for instance, by printing the original pages at six lines to the inch instead of eight.

AN EXTRACT FROM THE SUBJECT INDEX

```
Artificial satellites
    CLOUD COMPOSITION DETERMINATION BY SATELLITE SENSING USING THE NIMBUS VI         T78-6162
      HIGH RESOLUTION INFRARED SOUNDER
    ERROR ANALYSES OF OPERATIONAL SATELLITE SOUNDINGS OF VERTICAL TEMPERATURE        T78-6163
      PROFILES (12 1 -30 9 1977)
    TECHNOLOGY REQUIREMENTS FOR ADVANCED EARTH ORBITAL TRANSPORTATION SYSTEMS        T78-6275
      VOL 3: SUMMARY REPORT - DUAL MODE PROPULSION
Atmospheric circulation
    GENERALIZED COMPUTER PROGRAM FOR PRIMITIVE-EQUATION MODELS (1 10 1976-           T78-6164
      30 5 1977)
    OCEAN THERMAL RESPONSE TO STRONG ATMOSPHERIC FORCING  1: CHARACTERISTICS        T78-6139
      OF FORCING EVENTS; 2: THE ROLE OF ONE-DIMENSIONAL PROCESSES
```

AN EXTRACT FROM THE AUTHOR INDEX

```
Ainscough, J. B.
    EUROPIA CERAMICS FOR USE AS FAST REACTOR NEUTRON ABSORBERS                       T78-6110
Allan,T
    INTEGRATING PRODUCT DESIGN AND MANUFACTURE BY COMPUTER                           T78-6244
Anable, W
    SMELTING CEMENT COPPER IN AN ELECTRIC-ARC FURNACE                                T78-5944
Anderson, J R
    THEORY OF THE ACQUISITION OF COGNITIVE SKILLS (January-December 1977)           T78-5746
Andrews, A
    REMOTE ION BEAM POSITION MEASUREMENT FROM RANDOM BEAM MODULATION                 T78-6006
Armstrong, R D
    DEGENERACY IN SPECIAL PURPOSE PRIMAL ALGORITHMS USED IN OBTAINING LEAST         T78-5766
      ABSOLUTE VALUE ESTIMATORS
```

Fig.29 Sample indexes from R & D abstracts – TRC 1979

4.18.4 Contents pages (see Figure 18) can be produced automatically once the main text of an incremental issue of the announcement journal has been printed. The print program can incorporate logic for storing a list of the main headings and sub-headings used in a particular issue, along with the numbers of the pages upon which they occur. The data are then formatted and printed out to provide the reader with an easy index to the contents of the journal. These contents pages are placed with the standard introductory material which guides the reader in his use of the publication and advises him on such matters as the method of obtaining the actual reports announced in the journal. They are inserted before the main text and must therefore have a separate page-numbering system such as i, ii, iii etc.

4.18.5 Print quality depends very much upon the choice of printing process and, for orthodox on-line printing, the choice of computer printer. These decisions depend, in turn, upon the volume of work and the resources available, but the output from the most modestly priced printing device can be optimised by using opaque plain white paper and by carrying out each important printing run with a well cleaned machine and a new carbon ribbon. Printing work of lesser significance can be scheduled to fill in the time until the printer is required for the next good-quality run. There is no need, for instance, to use a slow printer and take special precautions when carrying out SDI runs, although upper and lower case printing is highly desirable especially if abstracts are provided, and plain paper is generally preferred.

4.19 Production Scheduling

It is quite impracticable to lay down a universal scale for the measurement of throughput in computer-based STI systems, so that any performance figures which are quoted by working organisations can only be regarded as a rough guide. The basic production parameters will reflect the quantity of documents passing through the system in a year and the frequency of issues of the journal in which they are announced. Dividing the quantity of documents by the number of issues will provide an indication of the number of items of bibliographic information to be announced in each incremental issue which, when multiplied by the average number of characters per item, gives the basis for assessing the rate at which the data prep facility can be expected to work. When making these calculations, allowances must be made for such factors as irregularity in the rate of document flow and the inevitable computer down time and, because the reputation of an STI organisation will be based upon the dependability of the services which it provides, these allowances should be made as generous as possible. It can be assumed that a VDU keyboard operator will provide between five and six hours of useful input in a working day and, in that time, should produce 30,000 corrected key depressions, i.e. about 5,000 key depressions per hour. This figure takes account of an assumption that approximately one third of the operator's time is spent in reading-back text which has already been keyed in. Thus, in systems where the initial proof-reading and correction of data are carried out outside the data prep room, a higher rate of key depressions can be expected.

Typical production cycles are based upon monthly, semi-monthly or fortnightly incremental intervals, and each issue of the announcement journal should be ready for distribution two or three working days before the date printed on the cover. In drawing up the production schedules it is essential to ensure that the print shop which converts the camera-ready copy created by the computer into bound books has adequate capacity, and that the data base handling capability of the organisation for the production of regular SDI output is sufficient to meet the users' needs. The journal editor, the computer operations manager and the data prep supervisor should consult together to draw up the schedules, and it is generally incumbent upon the editor to monitor progress and to attempt to clear hold-ups in the work flow as they arise.

4.20 File Security

There is no need to retain incremental disc files of bibliographic data for long periods once an issue of the announcement journal has been published and the data base has been updated. It is sensible to retain the journal page masters for up to two years in case a reprint of any issue should be called for in that time, and a similar retention period for incremental magnetic tapes could provide a valuable back-up service to the data base computer. Three 'generations' of the data base will be required on magnetic tape, with each generation updated as a new increment is added to the direct access disc file on the data base computer.

Care should be taken to protect new data as they are being entered from the VDU terminals. Systems for handling text is this way generally incorporate facilities for accessing a common file from a number of terminals at the same time, a practice which implies considerable risk of file corruption. It is therefore advisable to make a back-up copy of each input file at least twice a day and to update the incremental file on disc at the close of each day's operations, making a copy of the file on magnetic tape at the same time. At least three generations of the incremental magnetic tape should be maintained. When an issue of the announcement journal is ready for printing it is usual to set up a disc file containing the indexes which are to be printed with the journal and, where accumulated indexes are being created, arrangements will have to be made to update the related accumulated magnetic tape files. Back-up copies of the 'per issue' index files should be held on magnetic tape with the text files, and three generations of the accumulated index files should be maintained.

4.21 The Handling of Security Classified Information

The security of files described in paragraph 4.20 refers to essential precautions against the loss of vital data due to file corruption which may be caused by mechanical breakdown or human fallibility. Aspects of the handling of data which carry a security classification and must therefore be protected against unauthorised access are dealt with in Section 10, Vol.IV of the Manual, and the impact of such data upon a computerised system will depend upon the degree of secrecy involved. Bibliographic details of documents of the highest orders of confidentiality might never be stored in a computer, whereas less sensitive material might be handled only by nominated members of staff and held on separate files with back-up copies stored in a special limited-access safe. Wherever possible a special version of the computer operating system should be kept for use with security classified information, and precautions must be taken to ensure that all data are erased from core store and the internal work areas at the end of each security run.

4.22 Housekeeping and Stock Control

The main work of an STI Centre is the recording of complete report references for information processing and retrieval purposes in large bibliographic files, but this can be greatly assisted by the recording of a minimal description of each accession on internal files for housekeeping and stock control purposes. A typical stock control system can be based upon three files, a Reports File, a Transactions File and a Customer File, and sample records are shown in Figures 30, 31 and 32.

Reports reaching the Centre can comprise new material or duplicate copies of reports already in the system or returned loans, and it is strongly argued that the most error-proof way of operating the initial recording phase of the stock control system is to maintain a file of handwritten document Movement Cards against which all incoming reports can be immediately checked. By keying a minimal record defining each new accession into a Reports File from a terminal at the point at which reports are received however, it should be possible to meet the requirements of this initial recording operation. It would then be necessary to provide software with rapid information retrieval facilities operating in an interactive mode. The three basic stock control files are linked so that information common to more than one of them need only be keyed in once and, since requests for documents are recorded in the Transactions File, the system can be readily used for invoicing and accounting purposes where reports are sold, and for the printing of reminders for the control of overdue loans. Depending upon the rate at which documents flow through the system, several days can elapse between the initial recording of a new accession and the completion of the preparations for its appearance in the

Field Number	*Contents*	*Length (Characters)*
1	Accession number	8
2	Hardcopy or Microfiche	2
3	Title	24
4	Author	15
5	Pagination	3
6	Stock balance	4
7	Stock sent out (Stock movements)	4
8	Originator	21
9	Availability	9
10	First linking address	2
11	Second linking address	2

Notes:

1. Economies in file size can be achieved by placing binary values in the numeric fields.

2. Fields 10 and 11 are used to link the Reports File with the Customer and Transactions Files.

Fig.30 A typical Reports File record

Field Number	*Contents*	*Length (Characters)*
1	Account number	5
2	Name and address	118
3	Total number of requests received	4
4	First linking address	2
5	Second linking address	2

Fig.31 A typical Customer File record

Field Number	*Contents*	*Length (Characters)*
1	Request number	5
2	Customer account number	5
3	Accession number	8
4	Quantity ordered	4
5	Date issued	6
6	Date due back (loan)	6
7	Date of request	6
8	First linking address	2
9	Second linking address	2

Notes:

1. Economies in file size can be achieved by placing binary values in the numeric fields.

2. The linking address fields are used to link the three Stock Control files together.

Fig.32 A typical Transactions File record

announcement journal. The amount of keying required to identify each document for stock control purposes is small, and it generally simplifies the system and reduces the risk of errors if a fresh start is made when the full particulars of each accession are ready for keying into the main bibliographic files.

The Reports File entries created at the initial recording stage will contain the basic information required for the Document Processing Sheet (Fig.16) which travels with each document to control the complete recording operation. Where operational conditions permit, it is useful to eliminate the risk of transcription errors by holding the stock of blank processing sheets on continuous stationery and then printing out the initial identification data for each document via the computer. It is then necessary to match up each partially completed processing sheet with its parent document before the recording procedure, including the analysis of the report for indexing purposes, can be completed.

4.23 Forms Design

Many of the forms associated with the control of computer-based systems will consist of simple headed lists printed entirely by the machine and providing statistical data concerning the throughput of documents, etc. Forms which display textual data derived from the computer but intended for use outside the computer area, such as the Document Processing Sheets, invoice forms and index cards are greatly improved by being pre-printed on continuous stationery of a suitable size, and there are certain basic rules which have to be observed when such forms are being designed.

It is generally necessary to allocate space for each data field on the basis of ten characters to the inch across the form and six lines to the inch along its length. Many kinds of computer printer require a margin between the edge of the paper and the first print position in any line, and precise positioning of the text, which is essential if the completed form is to be acceptable to the users, can only be achieved if the computer operator sets up the paper accurately in relation to datum lines marked on the bed of the printer. If required, the edge strips containing the sprocket holes can be removed by guillotine or by means of perforations inserted when the blank forms are being pre-printed. The use of coloured paper can be beneficial, for tests have shown that black print on yellow paper is more legible than black on white. The use of two colours of ink on any form is not recommeded since it is of limited value and can add 20% to the cost of the pre-printed stationery.

Where up to six or eight copies are required of any pages of computer output it is generally most economical to use multi-part paper with carbon inserts, although the rather expensive 'no carbon required' papers are more convenient to handle. In the case of multi-part pre-printed stationery, it is often advisable to use different colours for the various parts to simplify sorting when the sheets are separated or 'burst'. Three-part invoicing forms represent a good example of this technique if the various parts are made instantly recognisable by having the invoice itself printed on white paper, the despatch copy on, say, yellow and the accounting copy on pink or some other contrasting colour.

4.24 Computer Configurations

The GEC 4080 computer used for the processing of research report literature at the Defence Research Information Centre of the UK MOD is illustrated in Figure 33. This configuration is described fully in the paper by Hart and included in the Conference Proceedings referenced as Item 12 in the Bibliography.

The Technology Reports Centre of the UK Department of Industry has a Molecular 18 computer of similar size, manufactured by Business Computers Limited. The TRC machine has the following configuration:

Central processing unit (96k bytes)
Operating console
1 32 megabyte disc unit
1 6.4 megabyte disc unit
1 Magnetic tape unit
6 Visual display key stations (VDU terminals) with 1280 character screens
1 Daisy-wheel keyboard printer (45 char/sec)
1 200 lpm printer (upper and lower case)
1 75 char/sec printer
1 paper-tape reader (500 char/sec)
1 paper-tape punch (40 char/sec)

5. THE USE OF MINI-COMPUTERS IN THE PROCESSING OF RESEARCH REPORT LITERATURE

The development of modern computer technology tends towards extreme miniaturization with ever-increasing data storage capacities occupying an ever-decreasing space, so that it becomes difficult to differentiate between the mini-computer and the main frame machine of moderate capacity. The optimum use of large data bases of the kind described in Chapter three can require up to ten 100 megabyte disc units on-line at any one time and, if access is provided for remote terminals over telephone land-lines or by satellite, the computer installation must also be equipped with special communications control equipment. The role of the mini-computer in installations of this size becomes that of the intelligent terminal fulfilling data capture and related word processing functions as well as providing off-line printing and

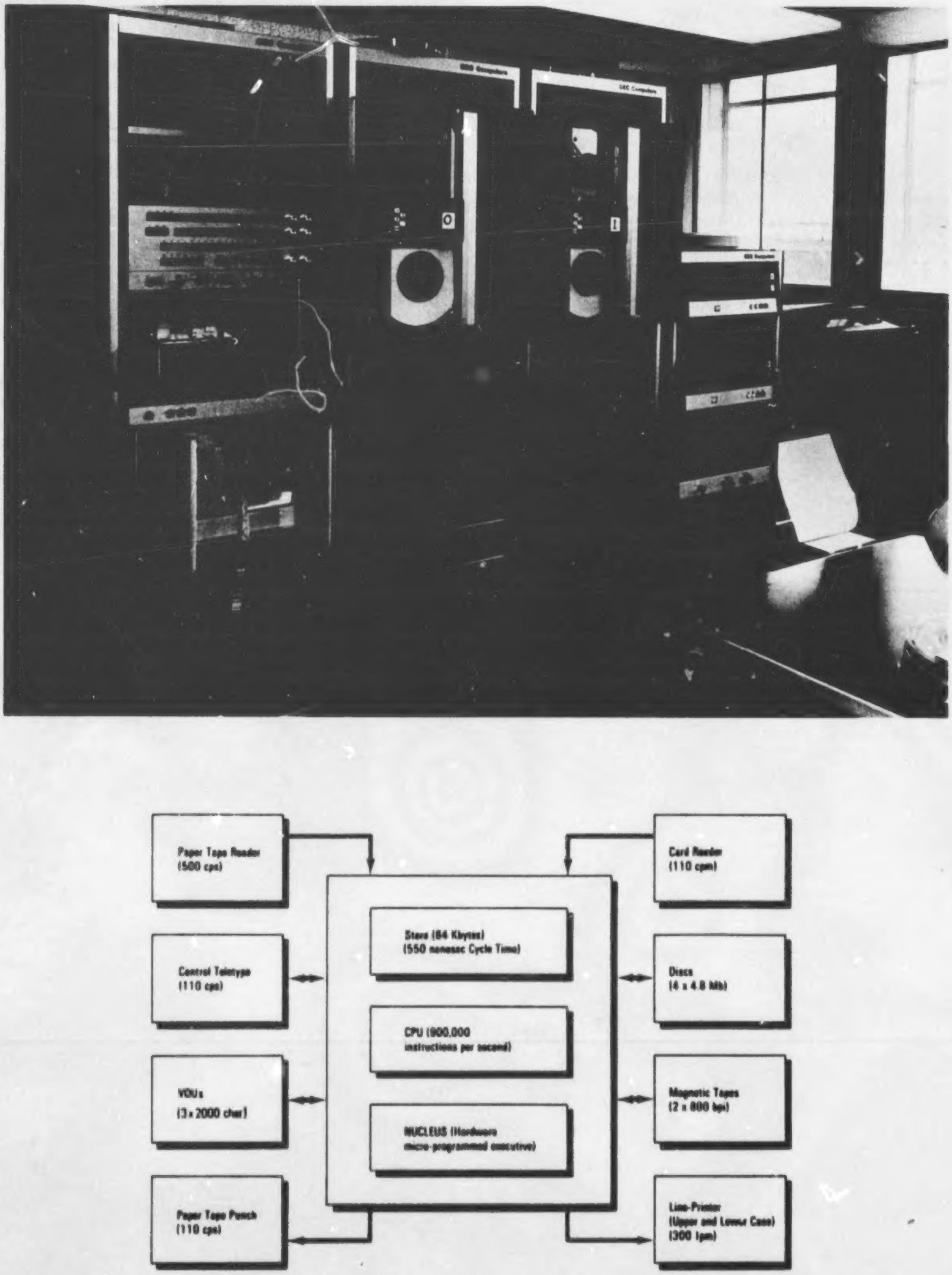

Paper Tape Reader
(500 cps)
Card Reader
(110 cpm)
Store (64 Kbytes)
(550 nanosec Cycle Time)
Control Teletype
(110 cps)
Discs
(4 x 4.8 Mb)
CPU (900,000
instructions per second)
VDUs
(3 x 2000 char)
Magnetic Tapes
(2 x 800 bpi)
NUCLEUS (Hardware
micro-programmed executive)
Paper Tape Punch
(110 cps)
Line-Printer
(Upper and Lower Case)
(300 lpm)

Fig.33 The GEC 4080 computer at DRIC
(by courtesy of the UK MOD)

communications control facilities. The communications control aspects of STI work are normally defined by the manufacturers of the main frame machine and the National Post Office authorities and form a highly specialised part of the total system. The use of mini-computers for word processing and as off-line print stations involves many of the techniques described in Chapter four and, if it is accepted that the term 'mini-computer' defines machines with core capacities up to 16k (16,000) bytes, the effect which this limitation will have upon file and program sizes must be considered when the total system is being designed.

5.1 Program Aspects

A mini computer equipped with VDU terminals and a simple printer can be used very effectively for on-line program development, with the programs being written onto floppy discs. The compilation of such programs, however, normally requires access to a much larger machine, and the sophisticated software offered by some manufacturers as an aid to error-tracing and the correction of programs (debugging) can rarely be accommodated in less than 32 k bytes of core if response times are to be acceptable to the programmers.

Mini-computers set up for word processing, i.e. data capture and the printing of good-quality output, can be enhanced to perform arithmetic, e.g. accounting, by the addition of suitable logic boards which, with the introduction of modern technology, do not require enlarged cabinets. There is, however, an inevitable constraint upon the number of programs which can be run simultaneously, and any over-loading of the system becomes particularly noticeable in the slowing-down of interactive operations, i.e. those activities which are based upon a dialogue between the user and the computer.

5.2 Non-bibliographic Applications

Reference has already been made in paragraph 4.22 to the value of stock control routines within a comprehensive STI system and, where the number of transactions and the quantity of material being handled justify it, these non-bibliographic functions can be usefully processed on a dedicated mini-computer. Access may be required to a main-frame machine, possibly the computer which holds the data base, either by a direct cable connection through an interface or by the use of compatible tapes, to run major sorting programs and to allow the updating of files which hold data common to both the stock control and main text handling systems. Once a stock control system has been set up in this way it can readily be extended to cover accounting procedures and various other housekeeping jobs. The mini-computer should be equipped with its own printer since it is a feature of some stock control functions that they require a dedicated printer for long periods although the total quantity of printing in the working day may be quite small. An example of this type of application is found in the initial recording of new accessions referred to in paragraph 4.22.

6. COMPUTER BUREAU OPERATIONS

There are many functions within a comprehensive STI system which require large computer resources. The accumulation of a single data base can result in the addition of 100 megabytes of on-line storage capacity each year; the large sorting programs can only work efficiently if more than 32k (32,000) bytes of main core are exclusively available while they are being run, and on-line information retrieval from ten remote terminals operating simultaneously can increase the demand for core capacity by up to eight times this figure, depending upon the efficiency of the software and the need for maintaining rapid response times between the computer and the terminals. Just printing out the results of searches in such a system could create a daily demand for over three hours exclusive use of a printer working at the rate of 400 lines per minute in upper and lower case and, although some components of computer hardware tend to become relatively less expensive as the technology develops, it is generally evident that installations which are large enough to cope with the demands of this kind of system can only operate economically if their resources are being fully exploited for the majority of a 24-hour day. The nature of the STI task however, creates peaks of utilisation during the normal working day with very light work-loads at other times.

One way out of this dilemma is to set up operations on the basis of distributed processing with the aid of a suitable computer bureau. This implies holding major files on a large main-frame machine linked, preferably by means of a dedicated telephone line, to terminals at the STI Centre. The bureau cannot be expected to exercise the kind of close control of files required for the systematic capture of reliable and accurate bibliographic data, neither would a general-purpose computer be normally equipped to cope with the special requirements of high-quality printing. Housekeeping and stock control activities such as those outlined in paragraph 4.22 might also prove impracticable without access to relatively simple local data storage and printing facilities, since parts of these files could be required for on-line access throughout the working day and a bureau may not be able to offer such facilities at an economic cost.

6.1 Setting Up the Bureau Operation

Two types of computer bureau can be recommended for use by an STI Centre. The usual kind is the purely commercial bureau which can only be expected to offer special facilities at a special rate of payment and may therefore prove to be very costly in the handling of large bibliographic files. The preferred type is the large-scale machine operated in a bureau mode by an educational or central government establishment, and organisations of this kind have been known to take a constructive interest in the special implications of STI processing and to help in overcoming the inevitable problems associated with large-scale text handling in a bureau environment.

Liaison with the bureau requires the attention of an expert in ADP on behalf of the Centre and, according to the scale of the operation, involves regular meetings with the bureau management particularly during the early stages of the operation. Users are expected to familiarise themselves with the standard procedures adopted by the bureau and to prepare batch work in accordance with any special rules which have been laid down. The conditions under which regular runs are carried out have to be agreed, with particular emphasis given to those jobs which are time-critical. It is often helpful for users to gain some understanding of the machine itself and its operating system in addition to a complete knowledge of the most efficient programming procedures.

An STI Centre which uses a bureau to handle its main files will select local word processing terminals for data preparation and peripheral devices for special printing etc. on the basis of obtaining the best value for the available money. It is likely then that there will be problems of compatibility with the main-frame machine which will generally provide its best response only if linked to terminals produced by specified manufacturers. These problems can be overcome by introducing software or hardware interfaces which will act as simulators, so that the bureau machine 'sees' the users' terminals, whatever their type, as devices which come within the recommended range of equipments.

6.2 The Advantages of Bureau Operation

The principal advantages of bureau operation have already been indicated, but it is also important that such a system should be shown to be cost effective over a period of seven to ten years. As well as offering the advantages of size, staff resources and a wide range of facilities, it is likely that the bureau will provide practical help and advice through some form of user support organisation. Thus progress towards a reliable working system can be achieved quite rapidly. The bureau operation will also provide the staff of the Centre with practical experience which could prove invaluable if it is decided, at a later date, to invest in a dedicated main-frame machine.

6.3 The Disadvantages of Burea Operation

The disadvantages of bureau operation are generally related to the fact that the STI Centre will inevitably relinquish absolute control over various important computer activities. This can be especially significant during the development period when the test runs of new programs are being submitted, since the time necessary for getting work into the computer and receiving results in the form of print-out can add days to the time required by the machine itself. It can be particularly frustrating when a job fails because of some simple error which would have been easily put right at run time on an in-house machine.

Operator errors at the bureau can also create problems, particularly if the printer is incorrectly set up, or the output is printed on the wrong paper or, upon completion, is sent to the wrong customer. It has been known for jobs to fail because the constituent programs have been run in the wrong sequence, and problems can also arise if output information is assigned to the wrong disc or magnetic tape file, thus corrupting essential data.

Despite these possible disadvantages, there is much to commend the use of a bureau to support computer-based STI operations. Many of the problems which have been described can be mitigated if user groups or similar liaison bodies are set up to disseminate information about system changes etc, possibly by the issue of news sheets. The local mini-computer linked to a remote main frame machine can, in the long term, form part of a communications network which might be developed to bring a number of STI Centres together for the mutually beneficial sharing of information resources.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the technical support provided by the UK Department of Industry and especially to thank Mr H.C.Adams of the Technology Reports Centre, Mr G.W.Hart of the Defence Research Information Centre and Mr Philip Bryant of the Bath University Library for assistance in the compilation and editing of the text.

BIBLIOGRAPHY

More detailed information concerning specialised aspects of the computer processing of scientific and technical information may be obtained from the following documents:

1.	Schipul, A.H.	*A Single Card Laboratory Reference Index System*: Army Medical Res. Lab., Fort Knox, Ky., USA: 240261 USAMRL-852, 18 February 1970: Inspec Report 71 CO 7535.
2.	Hammond, W.	*Construction of the NASA Thesaurus – Computer Processing Support* (final report): Aries Corp., Mclean, Va., USA: NASA-CR-95396:050168 NASW-1629: 5 January 1968: 68 N28811.
3.	–	UK MARC Manual – 1st Standard Edition – British Library Bibliographic Services Division (BLBSD) 1975 – ISBN 0-900220-47-3.
4.	Simpson, D.J.	*Before the Machines Come* – Aslib Proc. 20(1), pp.21-33, Jaunary 1968.
5.	–	*Documentation – Format for Bibliographic Information Interchange on Magnetic Tape*. International Standard – ISO 2709.
6.	–	*Bibliographic Information Interchange Format for Magnetic Tape*. British Standard BS 4748.
7.	Holloway, A.H. Ridler, E.H. Yates, B.	*Information Work with Unpublished Reports*: Institute of Information Scientists (UK) Monograph Series: 1976: André Deutsch.
8.	Metcalf, G.N. Evans, G.B.	*File Compression: A Report on the R & D Programme of the Birmingham Libraries Co-operative Mechanisation Project*. Birmingham, BLCMP 1977: (BLRDR 5359) 0 90315409 9.
9.	Spencer, H. Reynolds, L. Coe, B.	*The Relative Effectiveness of Ten Alternative Systems of Typographic Coding in Bibliographical Material*: 1973: Readability of Print Research Unit, Royal College of Art, 6A Cromwell Place, London, SW7 2NJ.
10.	Hyman, M. Wallis, E.	*Mini-computers and Bibliographic Information Retrieval*. Publications Dept. The British Library Lending Division, Boston Spa, Wetherby, West Yorkshire, LS23 7BQ. 5305HC: 1976: 0 85350 1440.
11.	Gillis, A.K.	*Methodology of Large Dynamic Files*: AGARD-R-649, 1976.
12.	–	*Advances in Retrieval Technology as Related to Information Systems*. AGARD Conference Proceedings No.207, October 1976.
13.	Gibbs, D.C.C.	*Glossary of Documentation Terms. Part II: Computer-user Terms*: AGARD-AG-182, Part II.
14.	–	*The Application of Inexpensive Minicomputers to Information Work*. AGARD Lecture Series No.92, March 1978.

GLOSSARY OF TERMS
and Acronyms

ADP	Automatic Data Processing
Batch processing	A mode of computer operation in which the data inputs for a number of tasks requiring the same program are accumulated over a period of time before being processed in a single machine run.
Bit	The smallest unit of information in a computer: a unit of storage capacity.
Block	A group of records, words, characters or digits handled as a unit in a computer system.
Boolean algebra	A binary system of algebra originally formulated by G.S.Boole and based on the operators AND, OR, NOT.
Buffer	Storage used to accommodate a difference in the data handling rate of two devices when information is being transmitted from one to the other.
Byte	A set of adjacent bits operated upon as a unit in a computer. A byte is generally smaller than a word and typically eight bits or six bits in length.
Character	A digit, letter or other symbol used in the representation of information. In computers a character is represented by a pattern of adjacent bits.
Character set	A finite set of different, specified characters.
Compile	To translate a source program into a program which the computer can execute.
Continuous stationery	A continuous strip of paper in roll or fanfold form, usually with uniformly spaced holes near the edges for sprocket drive purposes and a transverse line of perforations between sheets. Used as an output medium on line printers.
Control character	A character whose function is to initiate a computer operation.
COSATI	Committee on Scientific and Technical Information (see Section 1 of the Manual).
Data capture	The acquisition, in machine readable form, of data to be processed by a computer (also data preparation or data prep)
Down time	The time during which a computer is not available for operation due to malfunctioning.
Edit	To prepare data for subsequent processing or output by modifying the content, form or format.
Field	A specified part of a record reserved for a particular category of information.
File	A set of related records treated as a unit.
Flag	Any of various types of indicators used for identification: or a character that signals the occurrence of some condition such as the end of the word.
Floppy-disc *or* flexible disc *or* diskette	A low-cost, compact magnetic storage device comprising a small disc of flexible material. The information is stored on circular tracks in the conventional way, but each disc is contained in a sealed plastic envelope which has a slit from the centre to the outer edge. This slit provides access for the read/write head.
Hard-wired logic	A term used to describe the control of functions, within the computer or its peripheral equipment, by means of readily exchangeable printed circuit boards, generally as an alternative to the control of functions by program logic. The use of hard-wired logic is increasing with the development of microprocessor technology.
Interface	An area common to two systems or in which two systems interact with each other, e.g. the hardware linking a computer's central processing unit with the peripheral device.
Interpreter	A device that prints on a punched card the data already punched in the card.
Main frame	Synonymous with Central Processing Unit; generally implying a computer of large capacity.
NTIS	National Technical Information Service. Department of Commerce.
Off-line	Relating to the use of peripheral equipment, either uncontrolled by the central processing unit or initiated by the unit at a time subsequent to the current operation.

Parameter	A variable that is given a constant value for a specific purpose or process.
Record	A group of related items of information treated as a unit.
Run time	The time during which a computer program or routine is being executed.
SDI	The Selective Dissemination of Information.
Word	A set of adjacent bits or characters stored or operated upon as a unit in a computer.
Work area	A storage location reserved for temporary results.

Section 6

ANNOUNCEMENT SERVICES AND PUBLICATIONS

by
Elizabeth H.Ridler
Defence Research Information Centre
Ministry of Defence
Orpington, Kent, UK

ABSTRACT

The methods by which an Information Centre announces its holdings and recent acquisitions to its users are surveyed. Examples are given of manually and computer produced bulletins from a number of Information Centres and Libraries, and production of indexes to computer produced bulletins is discussed. Manual and computer-based SDI services are described with notes on profile construction, use of commercial magnetic tapes to extend an in-house SDI service, and standard profiles. Other services which involve repackaging the literature resources of a Centre are reviewed, including bibliographies, state of the art reports, and packaged information to assist industry with technical innovation.

CONTENTS

Page

1. INTRODUCTION 103

2. BULLETINS 103
   - 2.1 Literature Bulletins 103
   - 2.2 Accessions Bulletins 103
     - *2.2.1 Manually-Produced Bulletins* 104
     - *2.2.2 Computer-Produced Bulletins* 104
     - *2.2.3 Bulletin Indexes* 106
   - 2.3 Staff Requirements 107

3. SELECTIVE DISSEMINATION OF INFORMATION (SDI) 107
   - 3.1 The SDI Profile 109
   - 3.2 Definitions 109
   - 3.3 Manual Selective Dissemination of Information 109
   - 3.4 Computer-based SDI 110
     - *3.4.1 Compiling Profiles for the Computer* 110
     - *3.4.2 Example of Profile Compilation Using TEST Descriptors* 110
     - *3.4.3 Profile Compilation Using Natural Language* 112
     - *3.4.4 Other Methods of Profile Construction* 113
     - *3.4.5 Automatic Profile Construction* 113
   - 3.5 Output Assessment 113
   - 3.6 User Liaison 113
   - 3.7 Using Commercial Data Bases for In-house SDI 114
   - 3.8 Standard Profiles 114
   - 3.9 Supply of Documents from SDI Printouts 115
   - 3.10 Costs and Staffing Requirements of SDI 115

4. BIBLIOGRAPHIES 115

5. STATE OF THE ART REPORTS AND REVIEWS 116

6. PACKAGED TECHNOLOGICAL INFORMATION 116

7. NEWSLETTERS 117

8. THE FUTURE 117

LIST OF REFERENCES 117

Appendix 1 EUROPEAN SPACE AGENCY – ESA STANDARD PROFILES 120

107

1. INTRODUCTION

1.1 This section of the Manual deals with the types of announcement services and publications which are produced by an Information Centre as some of its major tasks. However extensive are the holdings of the Centre and valuable its acquisitions, they are of no real benefit until the Centre's users are made aware of their existence.

1.2 Information implies communication, a transfer of knowledge from a source to a receiver. Modern developments in using electronics for communications have not yet superseded the long-established medium of print on paper, and this section will deal mainly with the use of print on paper for communicating technical information in the form of regular literature and accessions bulletins, selective dissemination of information (SDI) services, special bibliographies, reviews, and other methods of packaging information. The information which is to be communicated will be mainly that derived by the information scientists of the Centre from original work already commited to paper by scientists and engineers in the form of books, technical reports, and papers given at conferences or published in learned and technical journals, the so-called 'primary publications'.

1.3 The introduction of computer processing in Information Centres has necessitated a degree of standardisation, but this does not mean that there is no room for individual styles in the publications of a Centre. As Kipling[1] has said: –

> 'There are nine and sixty ways of constructing tribal lays and every single one of them is right.'

Guidelines will therefore be given on methods for the production of the types of publication a Centre can provide with examples from services provided by national and other large information Centres to illustrate their different approaches to the problems. The examples will include some commercial services which provide information packages which can supplement an in-house service or provide models for the development of new services by the Centre.

2. BULLETINS

2.1 Literature Bulletins

Information Centres which serve their users with published information will prepare current awareness bulletins of articles from journals received in the Centre. In its simplest form the bulletin can be produced purely by clerical effort by circulating journal contents lists. A more useful bulletin is achieved if the information scientists in the Centre scan the journals and select the more important articles to announce. Although the marked up journals can be placed in the desired subject order and passed to a typist, the journals are out of circulation for a shorter period, and a more uniform product is obtained if bibliographical details of each selected article are transferred to a work sheet, on which the elements are arranged in the order in which they are to appear in the bulletin. A subject classification code is added for each item, the sheets arranged in subject order, subject heading slips intercollated, and with the addition of the title page the pack of work sheets is ready for typing and reproduction. Blick and Magrill[2] describe such a system for pharmacological research scientists. Information scientists scanned journals for items of interest to project teams. Titles of relevant items with author, bibliographical reference and language constituted the bulletin entry. A weekly bulletin containing 500–1000 references was circulated, saving each scientist one to two hours a week in journal scanning time. It was estimated that the cost benefit of the bulletin was equal to about 3.6 times its cost. This method of selection can approach manual selective dissemination of information services which are discussed more fully in paragraph 3.

The inclusion of abstracts in literature bulletins enhance their usefulness but also increases bulk and cost. There are, however, so many commercially available abstracting services that to prepare a literature bulletin with abstracts in an Information Centre is justified only if the Centre specialises in a particular discipline or an interdisciplinary subject, as many Information Analysis Centres do, or if the language used by the commercial abstract journals in the subjects required is not the native language of the Centre and the majority of its users. The basic steps for compiling a literature bulletin with abstracts will be the same as for a titles only bulletin, but an editorial staff will be required to maintain quality control of the abstracts and ensure a consistent product. The cost of the bulletin can be recovered by placing it on sale and two bulletins, not in the English language, are good examples of this course of action. In France the Centre National de la Récherche Scientifique (CNRS) issues the Bulletin Signalétique in 50 subject sections under the main headings Exact Sciences, Biological and Medical Sciences, Engineering Sciences. The bulletins are computer-produced and each is issued ten times a year, each issue containing subject and author indexes. (Further information from: Informascience, Centre de Documentation Scientifique et Technique, Centre National de la Récherche Scientifique, 26 Rue Boyer, 75971 Paris Cedex 20, France). In the Soviet Union the All-Union Institute for Scientific and Technical Information (VINITI) publishes Referativny Zhurnal, issued in several sections including Astronomy and Geodesy, Biochemistry, Biology, Chemistry, Electrotechnology, Geography, Geology, Geophysics, Informatics, Machine Construction, Mathematics, Mechanics, Metallurgy, Physics. Abstracting services on such a scale are usually computer-based, and produce computer searchable data bases as a by-product of bulletin production. Thus the PASCAL data base, available on the European Space Agency's Information Retrieval Service RECON system, is produced as a by-product of Bulletin Signalétique.

2.2 Accessions Bulletins

The accessions bulletin is one of the most important announcement media of an Information Centre, especially where the accessions relate to documents of limited availability such as technical reports where only a limited number are printed and the Centre may be the main distribution point for them. The function of such a bulletin is first to announce to users what has recently been received in the Centre, but a well-produced and indexed bulletin is also a permanent reference tool for retrospective search. The function of the bulletin in promoting the image of the Centre must also be borne in mind when considering its content and appearance.

The layout of the bulletin can be a matter of choice for the Centre, and some typical examples of bulletin entries will be illustrated in the following paragraphs. It will depend to a large extent on whether the bulletin is produced by manual methods or by computer. There are some general points which should be remembered. The user will scan each new issue to inform himself of new documents in his speciality. Items should therefore be grouped according to subject for his convenience. If the Centre uses a numerical classification such as the Universal Decimal Classification (UDC) the items are grouped under broad UDC headings. Centres which use subject indexing terms (descriptors) e.g. from the Thesaurus of Engineering and Scientific Terms (TEST)[3] will probably prefer to use a published set of subject category codes such as those issued by COSATI (see this Manual Vol.I, pp 14-15). Individual items should be laid out so that important information is prominent. Titles should be printed in capitals or in bold type, and the number which the Centre uses to identify the item in its system, usually the Centre's accession number, should be easily identified. It is helpful to users if introductory pages are included in the bulletin giving information on how to obtain the documents cited, and a guide to the layout of items. An example of such a guide, from the UK department of Industry's Technology Reports Centre's R & D Abstracts, is shown in Figure 1.

If some of the items to be announced are restricted to certain users because of security or commercial limitations, special supplements to the bulletin or a separate restricted edition should be compiled and circulated only to those users entitled to receive the information.

2.2.1 *Manually-Produced Bulletins*

The format of a manually-produced accessions bulletin will depend on the effort available in the Centre, and the use to which it is put. For immediate current awareness where speed of issue is all-important, a titles-only bulletin will be quick to produce and will involve the minimum of intellectual effort. It can be a simple typing operation based on copying marked fields on the work sheet made out for each new document, the sheets first being sorted into the subject order required. An example of such a bulletin is that currently issued by the British Library Lending Division, Boston Spa, as the BLLD Announcement Bulletin. A typical entry is shown in Figure 2. The entries are arranged in main COSATI subject fields. An author index is included.

A bulletin will be more informative if abstracts are included, although production will take longer, more information scientists' time will be involved and the typing and reprographic load will be greater. As with literature bulletins containing abstracts, an editorial staff will be required to ensure quality and consistent entries. Some economy can be achieved by coordinating bulletin production with other processes in the Centre. A Centre which does not have access to a computer for searching their records will probably rely on typed catalogue cards for its subject, author and other indexes. Bulletin and catalogue card production can be coordinated in various ways. After typing the cards one set can be arranged in bulletin page order, and offset litho plates made photographically. An alternative method is to type the bulletin pages first and to carry out two separate print runs, the first to prepare the bulletin pages and the second to prepare cards. Extra copies of each page can be run off on gummed paper, the abstracts cut out and stuck on to card. Alternatively the extra pages are run off on card stock which is subsequently guillotined to size. This latter method was used in the predecessor of the UK Defence Research Information Centre, the Ministry of Aviation's Technical Information and Library Service up to 1968 (Vessey[4]). These methods entail the use of a unit card for all the catalogues used in the Centre, although typed additions to the cards can be made if necessary.

In the intermediate stage before computerisation, tape typewriters, described in detail in Section 4 of this Manual, have been used to produce bulletins and cards from one typing operation. The cards are typed first and the bulletin is subsequently typed automatically using the punched tape produced when typing the cards. Wilson[5] describes the use of a Friden Flexowriter with edge-punched cards instead of paper tape at the Atomic Energy Research Establishment, Harwell, to produce catalogue cards, the monthly library accessions list and lists of unclassified UK Atomic Energy Authority reports from one master typing, with a saving in clerical and typing effort estimated at 1000 man hours a year. Vickers[6] describes the use of Vonamatic tape typewriters at the UK Central Electricity Generating Board Information Service to produce library catalogue cards, reports lists and a library bulletin, with considerable saving in typing and clerical effort.

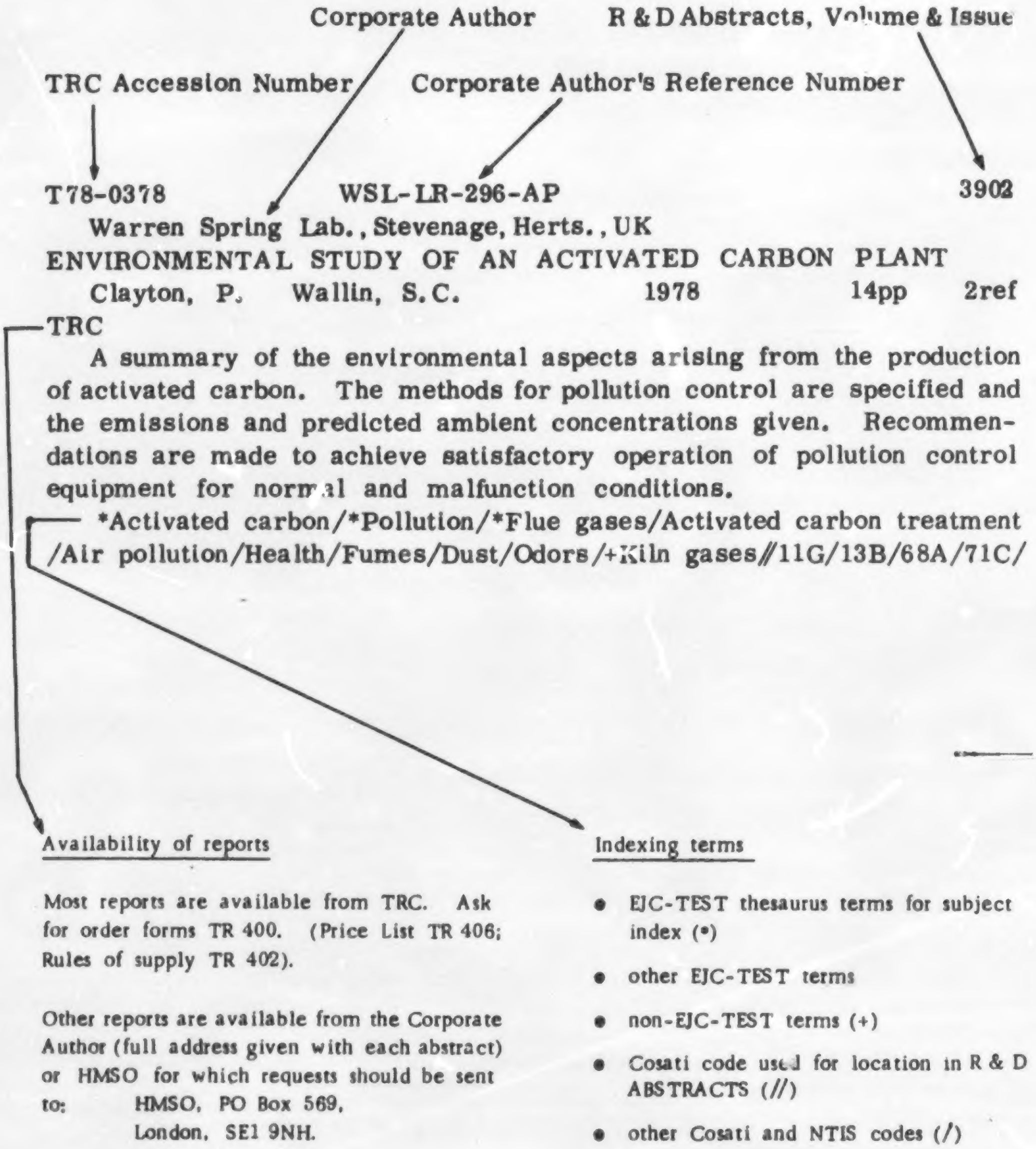
The main disadvantage with a manually produced bulletin is the difficulty of providing indexes, particularly subject indexes, which are the most important if the bulletin is to be used as a retrospective search tool. Extra cards must be written or typed and sorted into order and the final index retyped. For production of indexed bulletins some measure of computer assistance is essential.

2.2.2 *Computer-Produced Bulletins*

Data preparation for computer processing and computers in Information Centres are dealt with in detail in Sections 4 and 5 of this Manual. Only an outline of some actual systems used for bulletin production is therefore given here.

A TYPICAL ABSTRACT

BIBLIOGRAPHIC INFORMATION, ABSTRACT, AND INDEXING TERMS



Corporate Author

R & D Abstracts, Volume & Issue

TRC Accession Number

Corporate Author's Reference Number

T78-0378 WSL-LR-296-AP 3902

Warren Spring Lab., Stevenage, Herts., UK

ENVIRONMENTAL STUDY OF AN ACTIVATED CARBON PLANT

Clayton, P. Wallin, S.C. 1978 14pp 2ref

TRC

A summary of the environmental aspects arising from the production of activated carbon. The methods for pollution control are specified and the emissions and predicted ambient concentrations given. Recommendations are made to achieve satisfactory operation of pollution control equipment for normal and malfunction conditions.

*Activated carbon/*Pollution/*Flue gases/Activated carbon treatment /Air pollution/Health/Fumes/Dust/Odors/+Kiln gases//11G/13B/68A/71C/

Availability of reports

Most reports are available from TRC. Ask for order forms TR 400. (Price List TR 406; Rules of supply TR 402).

Other reports are available from the Corporate Author (full address given with each abstract) or HMSO for which requests should be sent to:

HMSO, PO Box 569,
London, SE1 9NH.

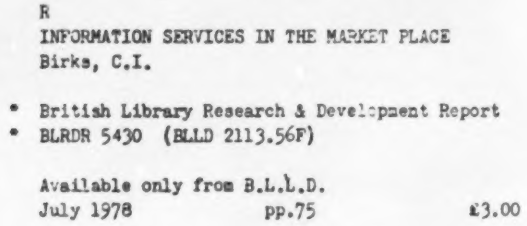
Indexing terms

- EJC-TEST thesaurus terms for subject index (•)
- other EJC-TEST terms
- non-EJC-TEST terms (+)
- Cosati code used for location in R & D ABSTRACTS (//)
- other Cosati and NTIS codes (/)

© Crown Copyright 1979

Extracts may be freely published providing the source is acknowledged.

Fig.1 Guide to lay-out of Technology Reports Centre's R & D Abstracts



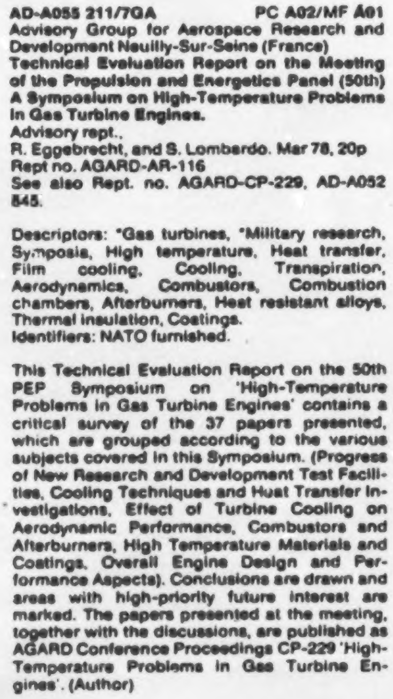
R
INFORMATION SERVICES IN THE MARKET PLACE
Birks, C.I.

- British Library Research & Development Report
- BLRDR 5430 (BLLD 2113.56F)

Available only from B.L.L.D.
July 1978 pp.75 £3.00

Fig.2 Typical entry in BLLD Announcement Bulletin

When a bulletin is produced by computer the bibliographical data must be prepared in machine readable form. Tape typewriters are frequently used for this purpose. Operational systems using tape typewriters have included the UK Technology Reports Centre's original system for R & D Abstracts (Schuler[7]) and the French CNRS system used for preparing Bulletin Signalétique (d'Olier and Dusoulier[8]). Other methods for computer input include Optical Character Recognition (OCR) used (1974) in the US Department of Navy Ship Systems and Scientific Documentation Division (Smith[9]) and direct entry into the computer from an on-line terminal using a visual display unit, as is now used for R & D Abstracts (Adams[10]). Advantages and disadvantages of each of these methods have been summarised by McIvor[11]. Development of minicomputers has led to the possibility of smaller Centres having their own computer, and bulletin production is one of the tasks very suitable for running on such machines. The preparation of the UK Defence Research Information Centre's abstract bulletin Defence Research Abstracts is now carried out completely in-house using a GEC 4080, the processes being described in detail by Hart[12] and in Canada the Defence Scientific Information Service prepares Documents Digest using a Varian 620i (McIvor[13]). The ultimate in computer production of bulletins is computer typesetting, described in Section 5 of this Manual. At present this is only economic for larger organisations such as NTIS and NASA. Figures 3 and 4 show typical entries in the NTIS Government Reports Announcements and the NASA STAR (Scientific and Technical Aerospace Reports).



AD-A055 211/7GA PC A02/MF A01
Advisory Group for Aerospace Research and Development Neuilly-Sur-Seine (France)
Technical Evaluation Report on the Meeting of the Propulsion and Energetics Panel (50th) A Symposium on High-Temperature Problems in Gas Turbine Engines.
Advisory rept.,
R. Eggebrecht, and S. Lombardo. Mar 78, 20p
Rept no. AGARD-AR-116
See also Rept. no. AGARD-CP-229, AD-A052 845.

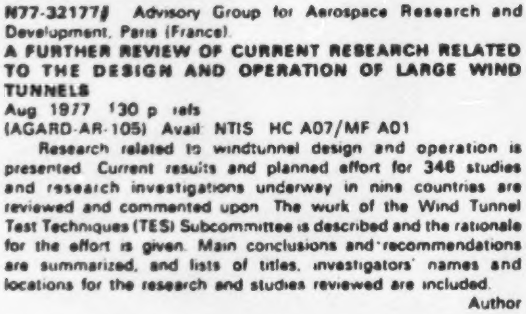
Descriptors: *Gas turbines, *Military research, Symposia, High temperature, Heat transfer, Film cooling, Cooling, Transpiration, Aerodynamics, Combustors, Combustion chambers, Afterburners, Heat resistant alloys, Thermal insulation, Coatings.
Identifiers: NATO furnished.

This Technical Evaluation Report on the 50th PEP Symposium on 'High-Temperature Problems in Gas Turbine Engines' contains a critical survey of the 37 papers presented, which are grouped according to the various subjects covered in this Symposium. (Progress of New Research and Development Test Facilities, Cooling Techniques and Huat Transfer Investigations, Effect of Turbine Cooling on Aerodynamic Performance, Combustors and Afterburners, High Temperature Materials and Coatings, Overall Engine Design and Performance Aspects). Conclusions are drawn and areas with high-priority future interest are marked. The papers presented at the meeting, together with the discussions, are published as AGARD Conference Proceedings CP-229 'High-Temperature Problems in Gas Turbine Engines'. (Author)

Fig.3 Typical entry in Government Reports Announcements

2.2.3 Bulletin Indexes

A manually produced bulletin, as stated earlier, is unlikely to include indexes because of the large amount of labour involved. Computer-produced bulletins can be indexed according to any of the data fields chosen. The most important index for retrospective search is the subject index, but the Centre will receive requests for documents not only by subject,



N77-32177# Advisory Group for Aerospace Research and Development, Paris (France).
A FURTHER REVIEW OF CURRENT RESEARCH RELATED TO THE DESIGN AND OPERATION OF LARGE WIND TUNNELS
Aug 1977 130 p refs
(AGARD-AR-105) Avail: NTIS HC A07/MF A01

Research related to windtunnel design and operation is presented. Current results and planned effort for 348 studies and research investigations underway in nine countries are reviewed and commented upon. The work of the Wind Tunnel Test Techniques (TES) Subcommittee is described and the rationale for the effort is given. Main conclusions and recommendations are summarized, and lists of titles, investigators' names and locations for the research and studies reviewed are included.
Author

Fig.4 Typical entry in STAR

but by author, title, document number, corporate author, contract number, etc. Printed indexes for each of these data elements will provide a search facility for the Centre's clerical staff who deal with requests for specific documents, and allow the Centre to dispense with a number of separate card catalogues with consequent saving in the time taken for card filing. Except for the subject index the indexes can be produced using simple computer sort programs. Consistency of recording is important and the rules given in Section 2 of the Manual must be followed to ensure standardized entries in the indexes.

The subject index is not so straightforward. It will depend on the indexing system used in the bulletin. If the UDC is employed, it may appear to be difficult to sort by computer, but programs have been developed for the purpose by the Zentralstelle für maschinelle Dokumentation, Frankfurt am Main and described by Schneider and Koch[14]. Subject indexes based on descriptors can be organised in different ways. Single descriptors (single or multiple term concepts) can be identified separately to the computer in the data preparation (see Manual, Section 4) and each is printed out in turn associated with the document title and, if desired, the full bibliographical reference. This method has been developed at the UK Aircraft Research Association Ltd and is described by Barnett[15]. Complex subject headings can be obtained by associating two descriptors. The main heading is identified by marking the relevant descriptor with an asterisk, and the qualifying descriptor is separated from it by a comma. The association is made intellectually at the editorial stage. This method, used in the NTIS Government Reports Announcements is also employed at the Defence Research Information Centre and is illustrated in Figure 5 which shows one item from Defence Research Abstracts and Figure 6 which shows the position of this item in the subject index.

Mention should also be made of a form of subject indexing based on manipulating document titles, KWIC (KeyWord In Context) and KWOC (KeyWord Out of Context). There are several computer programs available for preparing KWIC and KWOC indexes. A list of unimportant words is made – the 'stop list'. All other title words are selected in turn by the computer and either printed centrally down the page in alphabetical order, with the remaining title words in order on each side (KWIC), or printed at the side of the page (KWOC) with the whole title repeated to the right. Each entry is associated with an identification code. The usefulness of KWIC and KWOC indexes depends, of course, on document titles being informative.

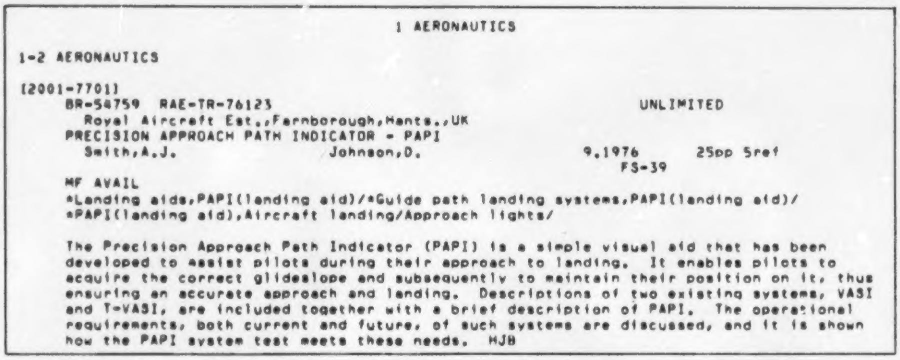
It is not necessary for every issue of the bulletin to contain a complete set of indexes. A useful minimum is subject, personal author, and an index associating the original document number with the Centre's accession number. A complete set of indexes should be issued as a separate volume or volumes, preferably at quarterly intervals with an annual cumulation. These index volumes will become the main search tools for the Centre for identifying individual documents and for conducting subject searches.

2.3 Staff Requirements

For a service where intellectual scanning of journals is required, but abstracts are not included, one information scientist can select, index and check about 300 citations per week if working full-time on the task[2]. When abstracts are included, a reasonable time for each abstract with indexing terms is 20–30 minutes, giving a weekly output of about 80 abstracts. If a large proportion of author abstracts can be used, 100 abstracts per week should be expected. The editorial staff can be expected to edit 300–400 items per week each, working full-time. Such staff would, however, normally have additional duties related to the Centre's other publications.

3. SELECTIVE DISSEMINATION OF INFORMATION (SDI)

As its name implies, SDI is a service in which information is tailored to user needs. Development of SDI services has increased rapidly with the increase in computer processing, although manual systems are successfully operated. Both manual and computer-based SDI will be considered, with the emphasis on the latter.



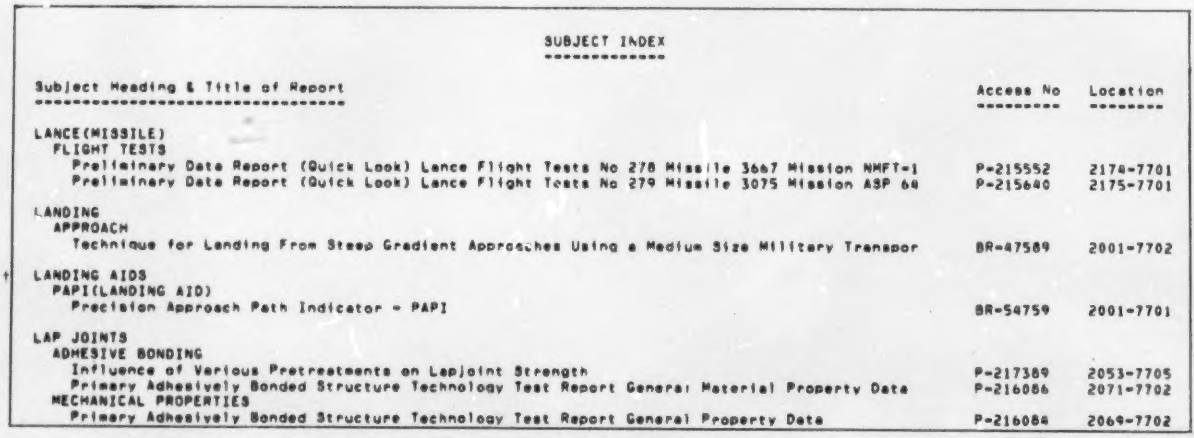
```
                              1 AERONAUTICS

1-2 AERONAUTICS

[2001-7701]
     BR-54759  RAE-TR-76123                                   UNLIMITED
       Royal Aircraft Est.,Farnborough,Hants.,UK
     PRECISION APPROACH PATH INDICATOR - PAPI
       Smith,A.J.             Johnson,D.                 9.1976      25pp 5ref
                                                            FS-39
     MF AVAIL
     *Landing aids,PAPI(landing aid)/*Guide path landing systems,PAPI(landing aid)/
     *PAPI(landing aid),Aircraft landing/Approach lights/

     The Precision Approach Path Indicator (PAPI) is a simple visual aid that has been
     developed to assist pilots during their approach to landing.  It enables pilots to
     acquire the correct glideslope and subsequently to maintain their position on it, thus
     ensuring an accurate approach and landing.  Descriptions of two existing systems, VASI
     and T-VASI, are included together with a brief description of PAPI.  The operational
     requirements, both current and future, of such systems are discussed, and it is shown
     how the PAPI system test meets these needs.  HJB
```

Fig.5 Extract from Defence Research Abstracts (DRA)



```
                                              SUBJECT INDEX
                                              =============

Subject Heading & Title of Report                                                                  Access No    Location
=================================                                                                  =========    ========

LANCE(MISSILE)
  FLIGHT TESTS
    Preliminary Data Report (Quick Look) Lance Flight Tests No 278 Missile 3667 Mission NMFT-1      P-215552     2174-7701
    Preliminary Data Report (Quick Look) Lance Flight Tests No 279 Missile 3075 Mission ASP 64      P-215640     2175-7701

LANDING
  APPROACH
    Technique for Landing From Steep Gradient Approaches Using a Medium Size Military Transpor      BR-47589     2001-7702

LANDING AIDS
  PAPI(LANDING AID)
    Precision Approach Path Indicator - PAPI                                                        BR-54759     2001-7701

LAP JOINTS
  ADHESIVE BONDING
    Influence of Various Pretreatments on Lapjoint Strength                                         P-217389     2053-7705
    Primary Adhesively Bonded Structure Technology Test Report General Material Property Data      P-216086     2071-7702
  MECHANICAL PROPERTIES
    Primary Adhesively Bonded Structure Technology Test Report General Property Data                P-216084     2069-7702
```

Fig.6 Extract from Cumulative Subject Index to DRA

3.1 The SDI Profile

The interests of the user must be translated into a form which can be used by the Information Centre to match against the material available so that only what is relevant is supplied. This is the user's 'profile'. When SDI is to be supplied to an individual scientist the first step in compiling his profile is to obtain a statement from him of his interests in his own words. This is facilitated if a standard form is used which should also include space for him to suggest keywords and to include references to known relevant documents. If the latter have been supplied by the Centre, the indexing terms assigned to them together with the keywords translated into the form used in the Centre's indexing system will give a preliminary list of terms from which the profile can be built up. For manual systems a fairly simple profile is sufficient. Computer processing allows of more sophistication, and a formal example of using descriptors from TEST for profile construction will be given in paragraph 3.4.2.

3.2 Definitions

Since SDI, especially when using a computer, involves information retrieval, definitions of a few commonly used terms are appropriate. These are natural language and controlled vocabulary indexing, and precision and recall in retrieved information. In *natural language (free text)* indexing, the indexing terms assigned are taken from the original text of the document. In *controlled vocabulary* indexing the indexing terms are taken from a fixed list of terms or thesaurus such as TEST. In assessment of retrieved output *precision* is the ratio

$$\frac{\text{number of relevant items retrieved}}{\text{total number of items retrieved}} \times 100$$

recall is the ratio

$$\frac{\text{number of relevant items retrieved}}{\text{total number of relevant items}} \times 100$$

3.3 Manual Selective Dissemination of Information

New documents received in an Information centre can be given an initial distribution to those most interested in the subject if a register of customers' interests is maintained. Such a register, often called a Field of Interest Register, normally consists of subject headings followed by abbreviations or codes indicating the organisations or individuals interested in the subject. It can be organised in different ways according to the method of subject indexing used in the Centre. A centre using TEST will probably find ordering by COSATI subject categories the most convenient, but numerical codes, e.g. the Universal Decimal Classification, or alphabetical subject headings, have also been used. An example of an Interest Register using alphabetical subject headings is given by Wright[16] who describes the Royal Aircraft Establishment Library's system. It consists of subject headings in alphabetical order with sub-headings followed by codes showing Departments and Divisions within Departments. It was originally prepared from a master index maintained on strip index panels from which new editions were produced by photography and offset printing, but its production has now been computerised.

The increasing use of microfiche enables new documents to be selectively distributed in this form, with consequent savings in waiting time and costs for postage and packing. A good example of such a service is that provided by NTIS in SRIM (Selected Research in Microfiche), in which an automatic distribution takes place every two weeks of microfiche reports in subject areas selected by the customer.

Manual SDI systems in which information about documents rather than the documents themselves is distributed are also feasible, although in many cases these are very similar to a Centre's literature bulletins already discussed in paragraph 2.1. Selection of items is often based on simple profiles constructed for groups of users interested in interdisciplinary subjects. Roysdon and Mistichelli[17] describe a system developed at Lehigh University for the interdisciplinary problems of Energy, Food, and Manpower/Womanpower. Abstracts on these subjects are copied from abstracting journals and primary sources by library staff to produce monthly continuing bibliographies. Circulation is to 50–80 recipients and the system is claimed to be cost effective and to have stimulated the use of library resources. Another manual system which involves building up individual profiles has been developed at the University of Aston, Birmingham, UK, and is described by Vincent and Seals[18]. The service is based mainly on Current Contents and other abstracting journals and serves ten groups of staff and research students in Pharmacy, Applied Psychology, Chemical Engineering, Management, and Production Engineering. References are copied on to cards, each card marked with profile code and library call number and a typed list prepared from the cards. Xerox copies of the lists are sent to the users.

The Scientific Documentation Centre, Ltd at Dunfermline, Scotland, operates a commercial manual SDI service. The Director, Dr P.S.Davison[19] describes the service as covering a broad subject field including spectroscopy, analytical chemistry, pollution, computer science, human and information sciences and librarianship. This involves scanning by part-time scientific staff annually 3250 journals, UK and US openly available government reports, UK and US thesis titles and 14000 book titles. A classification system devised by the Centre is used (Davidson[20]). Output is provided in the form of 125 x 75 mm cards distributed weekly, containing bibliographical details but not abstracts.

3.4 Computer-based SDI

Information retrieval using computer systems is dealt with in detail in Section 7, Vol.III of this Manual. It is sufficient to mention here that an SDI computer program will usually be very similar to the information retrieval program used by the Centre with possible modifications to the print program. The data base searched will normally be the latest update to the main data base, containing the input from one or two issues of the Centre's announcement bulletin, but it is possible to obtain other data bases by purchase or lease to enhance the service, and this possibility will be discussed in paragraph 3.7. A monthly SDI service is frequently provided. A service at longer intervals defeats the objective of immediacy, and unless the Centre has a very large input, or uses externally-produced tapes to supplement the service, a fortnightly SDI will produce only a few references.

3.4.1 Compiling Profiles for the Computer

For computer searching the search profile derived from the user's interest statement must be an ordered list of terms in a form strictly compatible with the indexing language of the data base to be searched and the search strategy used by the computer search program. The terms (descriptors) are sometimes all single words, as in many natural language systems or, as in TEST, single words and words combined to form concepts. Individual descriptors must, however, be coordinated to match the more complicated concepts of the user's interest statement. The method of coordination frequently used is the employment of Boolean logic operators. There are three of these operators commonly used, viz:

Operator	*Symbol*	*Example of use*	*Meaning*
AND	*	(1 * 2)	Select only documents that are indexed under both term 1 and term 2
OR	+	(1 + 2)	Select documents indexed under either term 1 or term 2
NOT	–	(1 – 2)	Select document indexed under term 1 but discard any indexed also under term 2

The processes can also be shown in the form of Venn diagrams as illustrated in Figure 7 where the circles represent sets of documents indexed by descriptors A and B respectively. The shaded areas represent (a) documents indexed both by A and by B, i.e. A * B, (b) documents indexed either by A or by B, i.e. A + B, (c) documents indexed by A but not by B, i.e. A – B.

3.4.2 Example of Profile Compilation using TEST Descriptors

As an example of profile compilation using controlled language, TEST descriptors in this case, the following steps can be recommended. Information is received from the user as follows:

Statement of interest	Use of carbon or glass fibre composite materials in aircraft structures
Relevant document(s)	Impact of composite materials on aerospace vehicles and propulsion systems. AGARD Structures and Materials Panel and Propulsion and Energetics Panel Symposium. September 1972, AGARD-CP-112.

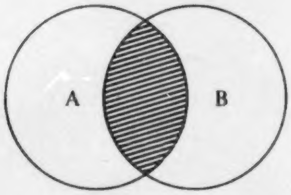
AGARD-CP-112 consists of a number of papers, and referring to the original indexing of these papers when processed in the Centre it is found that the following descriptors which appear relevant to the profile have been assigned:

Terms relevant to aircraft structures
Aircraft panels
Airframes

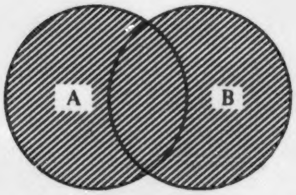
Terms relevant to composite materials
Composite fabrication
Composite materials
Composite structures
Fiber composites
Sandwich structures

Terms relevant to carbon or glass fibres
Carbon fibers
Fiberglass reinforced plastics

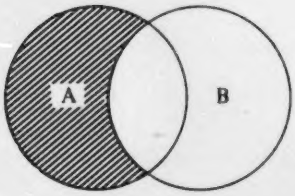
(It should be noted that the TEST spelling of descriptors is followed.) To make the profile more complete it is necessary to refer to these terms in TEST, and consider for selection some of the broader, narrower and related terms which are displayed. This results in selection of further terms as follows:



A AND B (A * B)



A OR B (A + B)



A NOT B (A − B)

Fig.7 Venn diagrams

Terms relating to aircraft structures
Control surfaces
Fuselages
Landing gear
Tail assemblies
Flaps (control surfaces)

Terms relating to composite materials
Chopped fiber composites
Oriented fiber composites

Terms relating to carbon or glass fibres
Fiberglass reinforced plastics

'Fiberglass reinforced plastics' defines a concept of a composite material using glass fibre. The same concept can be achieved by combining 'glass fibers' with the descriptors relating to composite materials using the AND operator. Using AND and OR operators the profile is built up line by line as follows:–

PROFILE

Line No.	*Term or logic equation*	*Notes*
1	Aircraft panels	Complete descriptor
2	Airframes	Complete descriptor
3	Control surfaces	Complete descriptor
4	Fuselages	Complete descriptor
5	Landing gear	Complete descriptor
6	Tail assemblies	Complete descriptor
7	Flaps (control surfaces)	Complete descriptor
8	Fins	Complete descriptor
9	1 − 8/+	1 OR 2 OR 3 OR . . . . . . OR 8

PROFILE

Line No.	*Term or logic equation*	*Notes*
10	Carbon fibers	Complete descriptor
11	Glass fibers	Complete descriptor
12	10 + 11	10 OR 11
13	Sandwich structures	Complete descriptor
14	Composite structures	Complete descriptor
15	Composite fabrication	Complete descriptor
16	Composite materials	Complete descriptor
17	Fiber composites	Complete descriptor
18	Chopped fiber composites	Complete descriptor
19	Oriented fiber composites	Complete descriptor
20	13 – 19/+	13 OR 14 OR 15 OR . . . OR 19
21	12 * 20	12 AND 20
22	Fiberglass reinforced plastics	Complete descriptor
23	21 + 22	21 OR 22
24	9 * 23 – PRINT	

The final line of the profile, 24, results in a printout of references meeting the requirements of the 'interest statement'.

Additional search facilities are usefully incorporated in an SDI computer program provided the data base is suitably structured so that data elements required to be searched are identified. Such additions include author names and COSATI fields. Each of these elements would be added to the profile and treated in the same way as the descriptors in the logic equations.

Economy in profile compilation can be achieved by truncation. A right hand truncation, represented often by a question mark, is exemplified by 'comput?'. Such a descriptor input would involve a search on all terms beginning 'comput' which in TEST are:

Computation
Computational linguistics
Computer components
Computer driven punches
Computerized simulation
Computer logic
Computer personnel
Computer programming
Computer programs
Computers
Computer storage devices
Computer systems hardware
Computer systems programs
Computer gun sights

By a similar process left hand truncation is represented by '?planes', which would bring out 'Airplanes', 'Rocket planes', 'Seaplanes'. The facility has to be used with caution as false drops can be obtained. In the 'comput?' example given above, for instance, 'computing gun sights' would be irrelevant in a profile designed to provide SDI on application of computers in information retrieval.

In controlled vocabulary indexing systems economy in profile construction is also obtained where the computer program used at the data input stage automatically posts documents indexed under specific, narrower terms to the broader, generic terms. In such systems the profile can consist of broader terms only, as the use of these terms will also retrieve the items indexed under the associated narrower terms. In the example profile given above, 'chopped fiber composites' and 'oriented fiber composites' are narrower terms both referred to 'fiber composites' as the broader term. If the 'posting up' procedure is followed at the input stage use of 'fiber composites' in the profile will retrieve documents indexed under 'chopped fiber composites' or 'oriented fiber composites', and these narrower terms can be omitted from the profile. The procedure will generate some 'noise', since documents indexed under any narrower term to a given broader term will be retrieved, and in some cases only a proportion of the narrower terms will be relevant to the profile.

3.4.3 Profile Compilation using Natural Language

A natural language profile can need more intellectual effort than a controlled vocabulary profile, as it is necessary to include all variations of a desired term in the profile. These variations include spelling, e.g. color/colour, computerise/computerize, sulphate/sulfate; and synonyms, e.g. columbium/niobium. For chemical subjects all alternative names for relevant chemical compounds must be included. A preliminary study of the vocabulary used in relevant documents is helpful in identifying synonyms. The profile usually consists of single terms, concepts being built up using the Boolean AND operator. Truncation is frequently employed, and other data fields, such as author names and classification codes are often included in the profiles. The main stages in profile construction are normally similar to the example 'TEST' profile given in paragraph 3.4.2.

3.4.4 Other Methods of Profile Construction

Although profiles are usually built up using words, it is possible to use numerical codes. Thus Brophy[21] describes a system developed at Teeside Polytechnic for informing library users of new acquisitions within their subject interests. The profiles are built up using Dewey Decimal Classification numbers.

The search strategy need not necessarily involve Boolean logic. The terms in the profile can be given weights (e.g. from 1 to 10) according to their relative importance to the search. The weights of all terms in the profile which match with terms in document records in the data base are summed to produce a document score. All document citations scoring above a given threshold level are printed out as the SDI output.

Evans[22] reports research on various search strategies as part of the INSPEC programme. Strategies ranged from simple coordinate matching of terms and groups of terms (concepts) to the use of term weights and/or group weights, and Boolean logic with and without weights. The best retrieval performances were given by strategies using weighting techniques and the most cost effective strategy overall was one which used the coordinate matching of a restricted list of terms with weights.

Farradane's system of relational indexing has been used experimentally for profile construction (Yates-Mercer[23]). Forty-three profiles were built up of concepts or groups of concepts linked by relational operators. Tested against a data base of 2829 abstracts from Metals Abstracts, the profiles gave an output having both precision and recall of 75%.

3.4.5 Automatic Profile Construction

Profile construction is labour intensive and uses highly trained staff. It is tempting to ask if profiles can be generated automatically by computer, or if computer assistance can cut the cost of intellectual effort. Some experimental work on these lines has been reported, although results to date do not lead to firm conclusions. Evans and Gould[24] report an investigation using the INSPEC thesaurus in which lists of natural language terms supplied by research workers were matched against the thesaurus. Three further profile versions were derived automatically by the addition in turn of (1) synonyms (2) narrower terms and (3) selected related terms. The profiles and their modifications were run against two document collections. The automatically generated profiles performed slightly worse than profiles compiled subjectively from the same set of user statements.

The United Kingdom Chemical Information Service (UKCIS) has carried out investigations using the free text file Chemical Abstracts Condensates (CAC). The work is reported in two papers (Barker et al. and Robson and Longman)[25,26]. The usefulness of a potential search term was quantified in terms of its specificity, a measure of the frequency of its occurrence in relevant material against its frequency of occurrence in the entire data base. Computer analysis of the text of relevant items from a fixed file of three issues of CAC was used to produce lists of terms in order of specificity, those above a given specificity value being added to the profile and the process repeated iteratively until no new relevant items were retrieved. The specificity lists were converted to standard search profiles which were compared against a different fixed file of CAC with conventionally produced profiles. The automatically produced profiles were cheaper to construct and performed similarly to the conventional profiles, although each type produced relevant items which the other missed.

3.5 Output Assessment

When the results of the first computer run of an SDI profile are submitted to the user, an assessment form should accompany the printout so that he can indicate the number of relevant, possibly relevant, and irrelevant references. These should be analysed by the compiler of the profile to investigate the reason for retrieval of irrelevant references. The data base searched should also be studied to find out if any relevant reports have been missed. The profile is modified if necessary and an assessment made of the modified profile. These steps should be repeated until the number of irrelevant references is reduced to an acceptable level. With SDI it is more important to have high recall than high precision in the response to the profile. A profile with high recall is likely to retrieve items which the user would not find for himself but which he will find of interest. A high-precision profile might only retrieve items already known to him.

3.6 User Liaison

For the best results from an SDI service the users should be encouraged to take an active interest. Information scientists in a Centre serving a local community such as a research establishment can be in frequent touch with the scientists receiving the SDI output. Such contacts can be on an informal basis, but a more formal approach is described by Butterly[27] of the National Nuclear Research Centre, Pretoria, South Africa, where structured interviews with users were employed in constructing and improving search profiles in an SDI system based on Atomindex tapes from the International Nuclear Information System (INIS) which uses a controlled language thesaurus.

Where an Information Centre serves a scattered community of users personal contacts may be more difficult. In such cases it is important to compile a guide to the system which is circulated to potential users. An example of such a guide is that published for the UK Department of Industry's Technology Reports Centre by Adams et al.[28]. This gives details of TRC indexing, use of TEST for descriptors, COSATI codes, and preparation of search profiles with examples. TRC profiles are based on the TRC data base and NTIS and INSPEC magnetic tapes. Burton[29,30] describes the system operated by the US Department of Agriculture's Technical Information Systems (formerly the National Agricultural Library). An SDI service serves over 1600 scientists from nine data bases. The users compile and modify their own profiles with the aid of a User's Guide and tutorial seminars. Another method of assisting users in large scale SDI services is employed in the Canadian CAN/SDI service run by the Canada Institute for Scientific and Technical Information, where local librarians and information scientists are trained as search editors[31].

3.7 Using Commercial Data Bases for In-house SDI

An SDI service based on the Centre's own input can be supplemented by running magnetic tapes of commercial data bases on the Centre's own computer. These tapes can be obtained by purchase or more usually by a leasing arrangement, sometimes with an additional royalty according to the number of items retrieved during an SDI run. Some programming effort will be required in the Centre to reformat the tapes to make them compatible with the Centre's search program, and SDI profiles may have to be rewritten. A profile written using a controlled vocabulary will need considerable modification if it is required to be matched against a free text file. Some examples of successful operation will show how the problems of using external data bases have been overcome.

In Australia the Aeronautical Research Laboratories introduced a computer-based SDI service for report literature in 1971, searching the data bases of the Australian Defence Science and Technology Information System (ADSATIS), and leasing the NTIS magnetic tape. ADSATIS acquisitions were already indexed using TEST descriptors. On the NTIS tapes not all descriptors used were from TEST, and a retroactive thesaurus was created and continuously updated. All descriptors on the ADSATIS tapes were searched, as were all asterisked descriptors on the NTIS tapes. A simple matching process using OR logic was employed, although this was extended later by using the facility of AND with any number of COSATI cateogries with each descriptor. Output, originally in the form of fortnightly individual computer printouts was later issued on continuous line printer perforated stationery, each item displayed being addressed to the recipient, the left card of each pair contining full bibliographical details, including abstract where available, and the right hand card being preaddressed to the library for loan request purposes. Sheppard[32,33] describes the development of the system for a PDP10 computer, six man-weeks of analysis and programming being required to start the service. After a few years experience a questionnaire completed by users showed that 96% of them considered the service to be of high or medium relative importance to them as a source of information.

Commercial data bases of published literature are also used by many Centres to provide SDI services to supplement or replace current awareness literature lists prepared in the Centre. The library of the UK Atomic Weapons Research Establishment has had many years experience in providing such a service. The development of the AWRE service has been described by Corbett[34,35,36] and Friend[37] starting with the Chemical Abstracts Service Chemical Titles magnetic tape, later supplemented by INSPEC and Nuclear Science Abstracts tapes. The computer program used was that developed by the National Research Council of Canada for the IBM 360/50 computer. Profiles, including truncated terms, are matched against title words. Author names and journal titles are used in some profiles, and subject codes also in INSPEC and NSA profiles. Benefits to the library have included reducing the necessity of multicopy purchasing of expensive primary and abstracting journals, and users' time in cover-to-cover scanning of journals has been reduced.

3.8 Standard Profiles

The production of personal profiles is a labour-intensive process, and the benefits of computer-based SDI can be obtained more cheaply by producing Standard Profiles. Standard Profiles are compiled by information scientists on topics of interest to groups rather than to individuals. Each profile is normally rather broader and simpler than a personal profile but the compilation process is similar. A number of information services produce both personal and standard profiles.

A good example of a Standard Profile service is that provided commercially by the European Space Agency Information Retrieval Service in its Standard Title series. These are issued monthly by computer search of the NASA/IAA data base. The titles available are shown in Appendx 1 (Information from ESA IRS, Via Galileo-Galilei, 00044 Frascati, Italy). In the United States the NASA SCAN (Selected Computer Aerospace Notices) service provides a similar twice-monthly product from this data base. Standard profiles based on the INSPEC data base are the INSPEC weekly Topics, covering published material in the fields of optics, quantum optics and particle optics, condensed matter physics, geophysics, devices and techniques in physics, power engineering, communications, electronics, computers and computing, control. (Information from: INSPEC, The Institution of Electrical Engineers, Savoy Place, London, WC2 0BL.) In France the Centre de Documentation Scientifique et Technique offer monthly Standard Profiles from the PASCAL data base. The profiles are compiled by information scientists in the Centre and will eventually cover 1000 titles. The output is computer printed four items to a page which is perforated so that individual items can be filed as 150 x 110 mm cards. Each item gives bibliographical details and includes an abstract. (Information from Informascience, Centre de Documentation Scientifique et Technique, Centre de la Recherche Scientifique, 27 rue Boyer, 75971 Paris Cedex 20.)

3.9 Supply of Documents from SDI Printouts

An SDI service based on the Information Centre's own data base will provide references to documents held in the Centre, and the SDI printouts can incorporate a request form for the user to mark up and return. When an SDI service also incorporates commercial data bases, the problem becomes complicated and the Centre's staff or the users' local librarians may find it time-consuming to locate the full text of some of the items cited on the commercial tapes. A partial solution to this problem is suggested by Bourne[38] at the University of California which has a geographically dispersed library system and runs an SDI service from tapes including CA Condensates, BIOSIS Previews, ERIC, and CAIN (AGRICOLA). A computer file was built up of journal title/location/call numbers. Output citation tapes from each SDI profile were run against the journal file as part of the output printing run and the library location was printed out after the citation. Such a system could be adapted to any Centre's needs if a union catalogue exists of journals held by libraries which cooperate in an inter-lending scheme with the Centre operating the SDI service.

3.10 Costs and Staffing Requirements of SDI

It will be clear from the preceding paragraphs that there are many factors affecting the cost of an SDI service. Whether external tapes are purchased or leased, or whether the service is provided solely from the Centre's own data base, the amount of work in profile construction expected from the user, and whether personal or standard profiles or both are provided are some of the main considerations. Guidance on the subject can best be given by indicating some published surveys which should be consulted for further details.

Information scientists' time in profile construction has been discussed by Evans[22] in the paper already cited in paragraph 3.4.4. The method used for profile construction has a considerable influence on the time taken. Compiling and modifying profiles using coordinate term matching averaged 90 minutes per profile. Using Boolean logic with weights the time was 168 minutes. Translated into yearly output an information scientist could handle about 1000 profiles using coordinate matching, or about 550 using Boolean logic with weights. Evans also found that profiles using a controlled vocabulary were compiled more quickly than those using free text. Hisinger[39] in describing the service provided by the National Technological Library of Denmark on profiles run against the COMPENDEX and INSPEC magnetic tapes estimates that profile construction takes 2–4 hours of an information scientist's time, with user cooperation.

Using external data bases introduces an extra cost. Kabi[40] of the United Kingdom Chemical Information Services (UKCIS) gives an analysis of costs (1972) of providing SDI from the CA Condensates magnetic tapes. Hisinger[39], cited above gives a detailed breakdown of (1971) costs which include a considerable element for procuring original material as 'information which is not accessible is of no practical use to the subscriber'. Multi-data base centres specialising in SDI have been reviewed by Zais[41] who suggests that costs tend to be distributed as follows: computer processing costs take approximately 30% of the budget, personnel costs about 60% and data base acquisition by purchase or lease about 10%. Vickers[42] has collected cost data from 18 operational computer-based systems in Europe and the USA using a structured cost analysis scheme. He tabulates SDI costs per user and unit costs of SDI processing. The figures show considerable variation.

4. BIBLIOGRAPHIES

The stimulus for preparing formal bibliographies can arise in various ways. An Information Centre may be asked to provide a bibliography before a research project is undertaken, to establish what work has already been carried out on the subject. The Centre may receive requests for literature searches on similar subjects from different customers which indicate a widespread interest in a new development. The information specialists in the Centre may consider that a development is of such importance that they should anticipate future demands for information and prepare a bibliography on the subject.

Before beginning the work it must be established what period is to be covered, and what type of publication is to be cited, such as openly published literature, unpublished reports, or both. The literature and reports collection of the Centre is the first source of the references to be included. If the subject is already included in the Centre's computerised SDI service, items can be cut from the printouts to provide the nucleus. Otherwise suitable abstracts from the Centre's announcement bulletin are copied. It is seldom possible, however, to complete a bibliography using in-house resources, and other available sources must be considered. Once some key papers are identified, the lists of references in them can be consulted and each citation used as a source for further retrospective search. The published Science Citation Index (Institute for Scientific Information, 325 Chesnut Street, Philadelphia, Pa, 19106, USA) in which journal articles citing a given reference are indexed, is helpful for this process. Centres with on-line access to systems such as Lockheed/DIALOG, can consult the ISI data base SCISEARCH®. An interactive on-line search to other data bases will produce additional references, and an output format which includes abstracts will give complete bibliographical entries without necessitating reference to the original source. For openly available report literature the data base provided by the US National Technical Information Service (NTIS) based on their Government Reports Announcements (GRA) is a good source, and NTIS information specialists themselves prepare bibliographies based both on the NTIS and on the Engineering Index data bases under the general title NTISearch. These bibliographies are advertised in GRA.

For aerospace report literature the NASA STAR/IAA data base is of particular importance, and this is available on-line via the NASA RECON (REmote CONsole) system in the USA and in Europe on the European Space Agency's Information Retrieval Service's RECON system. NASA issues a number of continuing bibliographies drawn from this data base which illustrate its value as a source for bibliographies. These include:–

Aeronautical Engineering (NASA SP-7037) monthly
Aerospace Medicine and Biology (NASA SP-7011) monthly
Earth Resources (NASA SP-7041) quarterly
Energy (NASA SP-7043) quarterly
Management (NASA SP-7500) annually.

They are available for purchase from NTIS.

The format of the bibliography must be carefully considered. The citations will normally be in the same format as those in the Centre's bulletin. Citations taken from other sources are best reformatted, and the whole collection retyped and reproduced by offset lithography. An extensive bibliography can usefully be grouped into subject sections and contain a separate author index. A shorter bibliography on a very specific subject can be arranged in author order. The title of the bibliography should indicate the period covered and, if limited to a particular type of publication, should indicate this in the title, i.e. 'Bibliography of report literature . . . .' or 'Bibliography of published information. . . . . . .'. Bibliographies should conform to the house style of the Information Centre and used in a special series with a distinctive cover.

5. STATE OF THE ART REPORTS AND REVIEWS

State of the art reports and reviews summarise recent progress in a specialised field but, unlike bibliographies, involve evaluation of the references cited. Their compilation begins with an initial survey of the relevant information in the same way as in comprehensive bibliographies, but the original papers must be obtained and read. Some expertise or at least background knowledge in the subject reviewed is an advantage in the compiler, and if this is not available in the Centre, a contract can be placed with a suitable expert, and the literature resources of the Centre placed at his disposal. Successful reviews can, however, be compiled by non-specialists and Geary[43] who has compiled a number of surveys issued by the British Scientific Instrument Research Association describes a method for use by scientists who are not specialists, relying mainly on abstracting journals for the primary search, leading to consultation of original papers, and discussions with specialists who are accessible. He points out that a non-specialist will acquire specialist knowledge during the search and can make a better job of writing for other non-specialists who will form the majority of his readers. He recommends that a review should be based on not less than 150 references, but should not exceed 400. Between six and twelve months should be allowed for its completion. The format of the review will consist of the text followed by a list of the cited references which preferably will include abstracts. Format of the review should conform to the house style of the Centre for technical reports.

6. PACKAGED TECHNOLOGICAL INFORMATION

Information packaged in particular ways is an important method of technological transfer, especially in assisting small firms who do not have information services of their own. In such cases information must be presented in a form tailored to the user's needs. A short abstract is seldom sufficient, and engineering assessments and diagrams can be included so that possible applications to the user's own production processes are made clear. The source of the information may be one or more technical reports, journal articles, or 'spin-off' from a defence or other government-sponsored programme which is considered by the sponsors to be suitable for unlimited dissemination outside the context of the original programme.

A good example of this type of service is the NASA Tech Brief. A collection of Tech Briefs is published quarterly, each issue containing a section on new product ideas, and individual Tech Briefs classified under the headings:– electronic components and circuits, electronic systems, physical sciences, materials, life sciences, mechanics, machinery fabrication technology, mathematics and information sciences. (Information from: Director, NASA Technology Utilisation Office, PO Box 875, Baltimore/Washington International Airport, Md 21240.) Individual Tech Briefs are also contributed to the NTIS Tech Notes Service. A folder of single page Notes, each illustrating the essentials of an item of applied technology, is issued every two weeks. Other contributors include the National Bureau of Standards, the Bureau of Mines, and the US Army, Navy and Air Force. Tech Notes are isssued in the subject categories: Computers, Electrotechnology, Energy, Engineering, Life Sciences, Machinery, Manufacturing Materials, Ordnance, Physical Sciences, Testing and Instrumentation. In the United Kingdom a rather similar service for UK industry was provided by the Department of Industry's Techlink (Schuler[44]). Each Techlink consisted of a one page digest, illustrated where necessary, dealing with a single topic or new technological development, with information on the source of follow-up information. It was classified under one or more of 40 subject categories which were used to effect selective distribution. The service in the form described was discontinued in 1978 and replaced by a TechAlert service in which Techlink type material is submitted for publication in technical journals with a wide circulation. The Technical Information Service of the National Research Council of Canada prepares Tech Briefs by analysing and selecting published material on technological innovation for Canadian manufacturing industry. Each Tech Brief is a digest of published literature of about five pages on a particular topic and has been described by Kirouac[45].

The preparation of such technological information packages requires considerable expertise, and knowledge both of the subject and of the customer's requirements. An engineering background and preferably experience in industry are important requirements for staff engaged in such work.

7. NEWSLETTERS

A Centre serving a large number of users covering a wide area should issue a periodic Newsletter, say at quarterly intervals. This can be a quite informal publication distributed to all users giving information on topics including new services proposed, staff changes in the Centre, and guidance on procedures such as requesting information or documents from the Centre. The main purpose of the Newsletter is to keep users in touch and to encourage them to make use of the services of the Centre in the most effective way. A Newsletter should be written with a light touch, and readability is an important quality. Projecting the image of the Centre as a friendly helpful place and not an impersonal department should be the aim at all times.

8. THE FUTURE

The increasing use of commercial data bases in on-line systems will affect many processes in Information Centres. Its main effect will be on the services which involve the exploitation of published literature in the Centre, because no commercially available data base will include controlled and security classified reports or other material of restricted distribution for which the Centre is responsible. As has been indicated, the commerical data bases can be used to augment the Centre's services, and provide the information which the Centre's information scientists will repackage. New methods for improving scientific communication will also affect information dissemination by the Centre. In this connection the work of the Capital Systems Group, Inc. sponsored by the US National Science Foundation on innovations in information transfer, should be noted. Creager[46] reports on the publication of a loose-leaf guidebook, half dealing with innovations and the remainder with creation of by-products, print-on-paper alternatives, non-print-on-paper and mixed media innovations, trends and prospects. Mention should also be made of a most interesting survey of computer assisted writing and editing systems compiled by Berman[47]. This indicates the possibilities of using the latest computer technology and suggests future trends in automated publishing systems.

LIST OF REFERENCES

1. Kipling, R. — *In the Neolithic Age. Stanza 5.*

2. Blick, A.R.
Magrill, D.S. — *The Value of a Weekly In-house Current Awareness Bulletin Serving Pharmaceutical Research Scientists.* Information Scientist 9, pp.19-28, 1975.

3. Engineers Joint Council — *Thesaurus of Engineering and Scientific Terms.* New York, 1968.

4. Vessey, H.F. — *Mechanization in Documentation. The Ministry of Aviation Technical Information and Library Services.* ASLIB Proceedings 16, pp.341-354, 1964.

5. Wilson, C.W.J. — *Use of the Friden Flexowriter in the Library of the Atomic Energy Research Establishment, Harwell.* Journal of Documentation 20, pp.16-24, 1964.

6. Vickers, P.H. — *Tape-typewriters; Practical Aspects of their Use in Information Work.* Information Scientist 1, pp.17-32, 1967.

7. Schuler, S.C. — *Mintech Reports Centre Services and Mechanization Programme.* Information Scientist 4, pp.47-64, 1970.

8. d'Olier, J.H.
Dusoulier, N. — *Bulletin Signaletique Mechanisation – the PASCAL System.* Program 5, pp.228-238, 1971.

9. Smith, R.C. — *SHARP: Experiences in Library Automation.* Special Libraries 65, pp.61-65, 1974.

10. Adams, H.C. — *Computer Processing of Bibliographic Data at the Technology Reports Centre of the Department of Industry.* TRC, 1977.

11. McIvor, R.A. *Use of Minicomputers in DSIS.* AGARD Lecture Series 92, pp.1.1-1.5, 1978.

12. Hart, G.W. *Production of an Abstracts Journal.* AGARD Lecture Series 92, pp.5.1-5.17, 1978.

13. McIvor, R.A. *A Minicomputer Based Information System.* AGARD Conference Proceedings 136, pp.2.1-2.7, 1973.

14. Schneider, K.; Koch, K-H. *The Use of the UDC in the Production of Mechanized Indexes.* Zentralstelle für Dokumentation, Frankfurt am Main. AMD-A-21, 1970. (English Edition of ZMD-A-10.)

15. Barnett, C. *A Computer-produced Keyword Indexing System for Technical Reports in the Library of the Aircraft Research Association Limited.* AGARD Conference Proceedings 136, p.1, 1973.

16. Wright, R.C. *Library and Information Services at the Royal Aircraft Establishment: Some Problems and their Present Solution.* AGARD Conference Proceedings 117, pp.2.1-2.17, 1972.

17. Roysdon, C.; Mistichelli, J. *Manual SDI, Information Services for Interdisciplinary Problem-oriented Research Groups.* Special Libraries 67, pp.358-366, 1976.

18. Vincent, L.; Seals, J.H. *A Manual Current Awareness Service at the University of Aston.* Aslib Proceedings, 27, pp.247-261, 1975.

19. Davison, P.S. *Operating Large-scale Broad Coverage CA and SDI Services.* Information Scientist, 6, pp.15-31, 1972.

20. Davison, P.S. *A Classification System Proposed for Current Awareness System Subjects.* SDC Bulletin No.4, pp.113-133, 1968.

21. Brophy, P. *Teeside Polytechnic Library's Acquisitions Information System.* Program, 11, pp.83-93, 1977.

22. Evans, L. *Search Strategy Variation in SDI Profiles.* INSPEC Rep.R75/21, p.129, 1975.

23. Yates-Mercer, P.A. *Relational Indexing Applied to the Selective Dissemination of Information.* J. Documentation, 32, pp182-197, 1976.

24. Evans, L.; Gould, A.M. *Automatic Aids to Profile Construction: Expansion of Search Terms by Thesaurus.* INSPEC Rep.R76/25, p.105, 1976.

25. Barker, F.H.; Veal, D.C.; Wyatt, B.K. *Towards Automatic Profile Construction.* J. Documentation, 28, pp.44-55, 1972.

26. Robson, A.; Longman, J.S. *Automatic Aids to Profile Construction.* J. American Society for Information Science, 27, pp.213-223, 1976.

27. Butterly, E. *Improving SDI Search Profiles.* Information Processing and Management, 11, pp.189-200, 1975.

28. Adams, H.C.; Rogers, V.J.; Ingledew, M. *Guide to Technology Reports Centre Selective Dissemination of Information Service.* Department of Industry, Technology Reports Centre, Orpington, Kent, BR5 3RE.

29. Burton, H.D. *A User-dependent SDI System.* Special Libraries, 64, pp.541-544, 1973.

30. Burton, H.D. *Educating SDI Users.* Special Libraries, 66, pp.252-255, 1975.

31. Gaffney, I.M. *CAN/SDI Current Status and Future.* Automation in Libraries, CACUL Workshop in Library Automation, pp.4.1-4.6, Vancouver, 1971.

32. Sheppard, M.O. *The Computer-based Selective Dissemination Service at Aeronautical Research Laboratories.* Department of Supply, Aeronautical Research Laboratories, Australia, Library Note 1, 1973.

33. Sheppard, M.O. *User Response to the SDI Service Developed at Aeronautical Research Laboratories, Australia.* AGARD-CP-179, pp.10.1-10.9, 1975.

34. Corbett, L. *Using Commercially Available Literature Tapes for a Current Awareness Service.* Information Scientist, 2, pp.83-102, 1968.

35. Corbett, L. *Progress at AWRE in the Development of Mechanised Current Awareness Services.* Program, 4, pp.141-149, 1970.

36. Corbett, L. *Problems in Using External Information Services – Attitudes of the Special Library and its Users.* Aslib Proceedings, 24, pp.96-110, 1972.

37. Friend, P.D. *The Use of External Data Bases to Extend Awareness Services based on Internal Resources at AWRE, Aldermaston.* Aslib Proceedings, 24, pp.678-685, 1972.

38. Bourne, C.P. *Improvements in the Coupling of SDI System Output with Document Delivery System.* J. Chemical Information and Computer Sciences, 16, pp.27-30, 1976.

39. Hisinger, B. *A Multidisciplinary and Computerised SDI Service for Industry and Research – Practical Experience and Costs.* Information Storage and Retrieval, 7, pp.175-187, 1971.

40. Kabi, A. *Use, Efficiency and Cost of External Information Services.* Aslib Proceedings, 24, pp.356-362, 1972.

41. Zais, H.W. *The Pricing of Information. A Model for Selective Dissemination of Information Services.* (PhD Thesis). California University, Lawrence Berkeley Laboratory, LPL-4257, 250p, 1975.

42. Vickers, P.H. *A Cost Survey of Mechanized Information Systems.* J. Documentation, 29, pp.258-280, 1973.

43. Geary, P.J. *Compilation of State-of-the-Art Surveys.* Bulletin of the Institute of Information Scientists, 3, pp.4-12, 1964.

44. Schuler, S.C. *Tailored Abstracts and Technical Digests. A Service of Industry.* AGARD-CP-117, pp.8.1-8.9, 1972.

45. Kirouac, G. *Technological Up-dating for the Manufacturing Industry.* AGARD-CP-179, pp.3.1-3.3, 1975.

46. Creager, W.A. *Innovations in Information Transfer: a Program to Stimulate Change.* AGARD-CP-225, pp.6.1-6.11, 1977.

47. Berman, P.I. *Survey of Computer assisted Writing and Editing Systems.* AGARD-AG229, 1977.

Appendix 1

EUROPEAN SPACE AGENCY – ESA STANDARD TITLES

233	Ablation and ablative materials
101	Adhesives
208	Aging of materials
214	Air and gas jets
248	Air traffic control and collision avoidance
221	Aircraft flight testing
247	Aircraft noise and sonic boom
102	Aircraft production and costs
183	Air cushion vehicles and Hovercraft
104	Aluminium applications
105	Antenna properties
283	Aurorae and airglow
107	Automatic flight control
109	Bearings and gears
244	Beryllium
299	Biometry and bioinstrumentation
111	Bonds, joints and fasteners
254	Boron
113	Carbon fibres
116	Ceramics
276	Cloud physics and seeding
117	Coatings and paints
228	Communication satellites
237	Communication systems for satellites, commands
205	Component and circuit reliability
118	Composite materials
117	Computer storage devices and memories
115	Computers, on board
120	Computerized design
119	Corrosion
212	Creep analysis and rupture strength
121	Cryogenics
122	Data acqusition
123	Dielectrics
280	Earth atmosphere
262	Earth resources: Agriculture
265	Earth resources: Data processing and sensors
263	Earth resources: Geology and geography
126	Elastomers
128	Electric batteries and fuel cells
114	Electronic display devices
223	Electric propulsion
124	Electronic packaging
282	Electrostatic probes
207	Environment simulation
129	Fibre technology
127	Fluidics
261	Flutter
257	Friction and wear in metals
274	Geodetic satellites and surveys
130	Graphite
268	Gravity gradient satellites
251	Holography
132	Hydraulic and pneumatic systems
131	Hydrology and hydrography
234	Inertial guidance and navigation
133	Information science
134	Infrared technology
246	Laminate, honeycomb and sandwich structures
241	Laser applications
135	Light sensing tubes
213	Linear and non-linear programming
225	Loads due to gusts/turbulence
136	Lubricants and lubrication
293	Lunar composition and geology
286	Magnetometers and magnetic probes
139	Management, business and administration
215	Management, projects
140	Materials testing
138	Metal forming
295	Meteorites and meteoroids
145	Microelectronics, design and fabrication
149	Non-destructive testing
148	Nuclear propulsion
151	Oceanography
152	Operations research
219	Optical telescopes
300	Orbit calculations
202	Outgassing/degassing
275	Ozone
154	Photochemistry
272	Piezoelectricity
155	Pipe flow
157	Plastics
255	Plate buckling

158 Pollution, air
142 Pollution, noise
141 Pollution, water
159 Polymers and polymerization
161 Powder metallurgy
164 Pressure measurements
167 Pulse code modulation

166 Quality control

169 Real time operation
256 Reflective coatings
229 Rocket propellants

230 Satellite attitude control
272 Seals
269 Self-adaptive control systems
177 Semiconductor diodes and rectifiers
176 Semiconductor materials
266 Solar cells
227 Spacecraft power supplies
180 Stainless steels
182 Superconductors

258 Temperature measuring instruments
184 Thermal insulation
185 Thermal protection
186 Thermosetting resins
238 Thin film properties
204 Time measurement
209 Titanium
232 Trajectory optimization
190 Transducers
178 Transistors
179 Transistor circuits
181 Transmission lines
191 Transport economics and management
177 Tropopause/Troposphere
192 Tube heat transfer

193 Vacuum technology
194 Ventilation and cooling
216 V/STOL aircraft

196 Water purification
146 Waveguides, resonators and microwave tubes
278 Weather forecasting
198 Welding

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-AG-235 Volume II	ISBN 92-835-1331-2	UNCLASSIFIED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title MANUAL OF DOCUMENTATION PRACTICES APPLICABLE TO DEFENCE-AEROSPACE SCIENTIFIC AND TECHNICAL INFORMATION – Volume II: Data Recording & Storage, Mechanization Systems & Operations, Announcement Services & Publications.

7. Presented at

8. Author(s)/Editor(s)	9. Date
S.C.Schuler, General Editor	July 1979

10. Author's/Editor's Address	11. Pages
	126

12. Distribution Statement This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Information sciences
Computing
Data recording
Data storage
Minicomputers
SDI
Libraries
Personnel
Publications

14. Abstract

The second of four separately published volumes describing the basic documentation practices involved in the initial setting up and operation of an Information-Library organisation to provide defence-aerospace information services. The focus is on a practical, rather than theoretical, approach for both the senior person setting up a new system as well as junior staff who may be using the manual as a training aid.

This volume consists of three main sections. The first two are concerned with basic mechanisation and the use of computers for information processing. An introduction to the hardware and software of computer systems is given. Detailed aspects such as input of data, storage devices, output printers, system design and bureau working are discussed. Some guidelines on the processing of chemical structure input is included. The third section describes the organisation and methods of production involved in various announcement publications and services including SDI.

This AGARDograph was prepared at the request of the Technical Information Panel of AGARD and publication of the remaining volumes will occur during the next two years.

AGARDograph No.235
Advisory Group for Aerospace Research and Development, NATO
MANUAL OF DOCUMENTATION PRACTICES APPLICABLE TO DEFENCE-AEROSPACE SCIENTIFIC AND TECHNICAL INFORMATION
Volume II – Data Recording and Storage, Mechanization Systems and Operations, Announcement Services and Publications
S.C.Schuler, General Editor
Published July 1979
126 pages

The second of four separately published volumes describing the basic documentation practices involved in the initial setting up and operation of an Information-Library organisation to provide defence-aerospace

P.T.O.

AGARD-AG-235 Vol.II

Information sciences
Computing
Data recording
Data storage
Minicomputers
SDI
Libraries
Personnel
Publications

AGARDograph No.235
Advisory Group for Aerospace Research and Development, NATO
MANUAL OF DOCUMENTATION PRACTICES APPLICABLE TO DEFENCE-AEROSPACE SCIENTIFIC AND TECHNICAL INFORMATION
Volume II – Data Recording and Storage, Mechanization Systems and Operations, Announcement Services and Publications
S.C.Schuler, General Editor
Published July 1979
126 pages

The second of four separately published volumes describing the basic documentation practices involved in the initial setting up and operation of an Information-Library organisation to provide defence-aerospace

P.T.O.

AGARD-AG-235 Vol.II

Information sciences
Computing
Data recording
Data storage
Minicoputers
SDI
Libraries
Personnel
Publications

AGARDograph No.235
Advisory Group for Aerospace Research and Development, NATO
MANUAL OF DOCUMENTATION PRACTICES APPLICABLE TO DEFENCE-AEROSPACE SCIENTIFIC AND TECHNICAL INFORMATION
Volume II – Data Recording and Storage, Mechanization Systems and Operations, Announcement Services and Publications
S.C.Schuler, General Editor
Published July 1979
126 pages

The second of four separately published volumes describing the basic documentation practices involved in the initial setting up and operation of an Information-Library organisation to provide defence-aerospace

P.T.O.

AGARD-AG-235 Vol.II

Information sciences
Computing
Data recording
Data storage
Minicomputers
SDI
Libraries
Personnel
Publications

AGARDograph No.235
Advisory Group for Aerospace Research and Development, NATO
MANUAL OF DOCUMENTATION PRACTICES APPLICABLE TO DEFENCE-AEROSPACE SCIENTIFIC AND TECHNICAL INFORMATION
Volume II – Data Recording and Storage, Mechanization Systems and Operations, Announcement Services and Publications
S.C.Schuler, General Editor
Published July 1979
126 pages

The second of four separately published volumes describing the basic documentation practices involved in the initial setting up and operation of an Information-Library organisation to provide defence-aerospace

P.T.O.

AGARD-AG-235 Vol.II

Information sciences
Computing
Data recording
Data storage
Minicomputers
SDI
Libraries
Personnel
Publications

information services. The focus is on a practical, rather than theoretical, approach for both the senior person setting up a new system, as well as junior staff who may be using the manual as a training aid.

This volume consists of three main sections. The first two are concerned with basic mechanisation and the use of computers for information processing. An introduction to the hardware and software of computer systems is given. Detailed aspects such as input of data, storage devices, output printers, system design and bureau working are discussed. Some guidelines on the processing of chemical structure input is included. The third section describes the organisation and methods of production involved in various announcement publications and services including SDI.

This AGARDograph was prepared at the request of the Technical Information Panel of AGARD and publication of the remaining volumes will occur during the next two years.

ISBN 92-835-1331-2

information services. The focus is on a practical, rather than theoretical, approach for both the senior person setting up a new system, as well as junior staff who may be using the manual as a training aid.

This volume consists of three main sections. The first two are concerned with basic mechanisation and the use of computers for information processing. An introduction to the hardware and software of computer systems is given. Detailed aspects such as input of data, storage devices, output printers, system design and bureau working are discussed. Some guidelines on the processing of chemical structure input is included. The third section describes the organisation and methods of production involved in various announcement publications and services including SDI.

This AGARDograph was prepared at the request of the Technical Information Panel of AGARD and publication of the remaining volumes will occur during the next two years.

ISBN 92-835-1331-2

information services. The focus is on a practical, rather than theoretical, approach for both the senior person setting up a new system, as well as junior staff who may be using the manual as a training aid.

This volume consists of three main sections. The first two are concerned with basic mechanisation and the use of computers for information processing. An introduction to the hardware and software of computer systems is given. Detailed aspects such as input of data, storage devices, output printers, system design and bureau working are discussed. Some guidelines on the processing of chemical structure input is included. The third section describes the organisation and methods of production involved in various announcement publications and services including SDI.

This AGARDograph was prepared at the request of the Technical Information Panel of AGARD and publication of the remaining volumes will occur during the next two years.

ISBN 92-835-1331-2

information services. The focus is on a practical, rather than theoretical, approach for both the senior person setting up a new system, as well as junior staff who may be using the manual as a training aid.

This volume consists of three main sections. The first two are concerned with basic mechanisation and the use of computers for information processing. An introduction to the hardware and software of computer systems is given. Detailed aspects such as input of data, storage devices, output printers, system design and bureau working are discussed. Some guidelines on the processing of chemical structure input is included. The third section describes the organisation and methods of production involved in various announcement publications and services including SDI.

This AGARDograph was prepared at the request of the Technical Information Panel of AGARD and publication of the remaining volumes will occur during the next two years.

ISBN 92-835-1331-2

AGARD
NATO OTAN
7 RUE ANCELLE · 92200 NEUILLY-SUR-SEINE
FRANCE
Telephone 745.08.10 · Telex 610176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD – VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Defence Scientific Information Service
Department of National Defence
Ottawa, Ontario K1A 0Z2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92 Châtillon sous Bagneux

GERMANY
Zentralstelle für Luft- und Raumfahrt-
dokumentation und -information
c/o Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens, Greece

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direcção do Serviço de Material
da Forca Aerea
Rua da Escola Politecnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy
National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

Microfiche
Space Documentation Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

Microfiche
Technology Reports
Centre (DTI)
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RF
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London W1P 1HD

ISBN 92-835-1331-2